Citrate synthase from the halophilic archaeon Haloferax volcanii

Maddocks, Deborah G.

Award date:
1999

Awarding institution:
University of Bath

Citrate Synthase from the Halophilic Archaeon *Haloferax volcanii*

Submitted by Deborah G. Maddocks
For the degree of PhD
Of the University of Bath
1999

COPYRIGHT

D. G. Maddocks.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

Haloferax volcanii is an extremely halophilic Archaeon that was originally isolated from hypersaline environments where the salt concentration approaches saturation (>5 M NaCl). In order to survive at such extreme conditions the organism maintains an intracellular concentration of salt, mainly KCl, that is isotonic with its environment. Therefore, enzymes from extreme halophiles such as *Hf volcanii* are fully active at salt concentrations up to 5 M and indeed they require these conditions in order to retain full activity. Thus enzymes from these organisms have potential biotechnological applications in both high salt environments and other situations of low water activity.

We have used citrate synthase as a model enzyme for an investigation of the structural basis of enzyme halophilicity. The initial aim of the investigation was to isolate, clone and sequence the citrate synthase gene from *Haloferax volcanii*. This has been achieved using a *Haloferax* genomic library, which was probed with a 880bp citrate synthase gene fragment obtained from a previous study.

The gene was 1138 nucleotides in length which translated into a 380 amino acid protein. A potential promotor was identified but further analysis was required to confirm that transcription originated from the predicted site. Sequence homology searches showed that the gene had the highest sequence identity to the Gram-positive Bacterium *Bacillus subtillis* closely followed by the Archaeon *Archaeoglobus fulgidus*. Phylogenetically the *Haloferax* citrate synthase gene grouped within the Gram-positive Bacterial/Archaeal branch of both phylogenetic trees produced.

Expression of the recombinant protein was achieved using both *Haloferax* and *E.coli* as host systems. Recombinant protein was expressed in the homologous form using the pBAP5009 halophilic shuttle vector; and in the heterologous form in the commercial pET vector system. The pET vector was selected over the pBAP vector as the expression system of choice, mainly due to the large amount of protein expressed by the system.

A structural model of the halophilic enzyme has been produced based on the crystal structure of citrate synthase from the hyperthermophilic Archaeon *Pyrococcus furiosus*. The model exhibits features typical of halophilic proteins, including a high density of negatively charged amino acids covering the surface of the protein. The contribution of the number and spatial arrangement of these residues to the halophilicity of the protein has been explored through site directed mutagenesis. A putative K^+ binding site was identified at the dimer interface of the enzyme, consisting of 4 co-ordinated residues, 2 glutamate and 2 aspartate; one of each contributing from each monomer. This site is apparently absent from the enzyme from non-halophilic organisms and was deemed a suitable target for site-directed mutagenesis.

The site was altered and mutant proteins expressed in *E.coli.* Characterisation of these mutants did not revealed any major changes in the halostability of the enzyme, however, a potential intra-molecular ionic network has been identified in the region of interest which could play a role in conferring thermal stability. Ionic networks have been seen in other organisms in particular hyperthermophiles where their role has been shown to be linked to maintaining stability of the enzyme at elevated temperatures.

Table of Contents

Page Number

Chapter 1: Introduction

Chapter 2: Materials and Methods

Chapter 3:
Cloning & sequencing of the Citrate Synthase Gene from Haloferax volcanii

Chapter 4:
Analysis of the Hf volcanii Citrate Synthase Gene

Chapter 5: Construction of a Molecular Model of Haloferaxvolcanii Citrate Synthase Based upon the Crystal Structure of the Enzyme from Pyrococcus furiosus

Chapter 6: Expression of Citrate Synthase in its Native Host Haloferaxvolcanii, and in the Gram-Negative Bacterium Escherichia coli

Chapter 7: Site-Directed Mutagenesis of Haloferaxvolcanii Citrate Synthase

Chapter 8: General Discussion

List of Figures

List of Tables

Abbreviations

ATP	adenosine triphosphate
bp	base pairs
BSA	bovine serum albumin
CoA	coenzyme A
CS	citrate synthase
DHLipDH	dihydrolipoamide dehydrogenase
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DTNB	5,5'-dithio bis-2-nitrobenzoate
EDTA	(disodium) ethylenediamine tetraacetate
Fd	ferredoxin
IPTG	isopropyl-β-D-thiogaalactosidase
kb	kilobase pairs
kD	kilodaltons
LB	Luria-Bertaini
MGM	modified growth media
MOPS	3-(N-morpholino) propane-sulfonic acid
M_r	molecular mass
NAD(H)	nicotinamide adenine dinucleotide (reduced)
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PECSS	Perkin Elmer computer spectroscopy software
PEG	polyethylene glycol
PMSF	phenylmethane sulfonyl fluroride
RNAse A	ribnuclease A
rRNA	ribosomal RNA

SDS	sodium dodecyl sulfate
TAE	tris-acetate EDTA
TE	10 mM Tris-HCl, 1mM EDTA (pH 8)
TEMED	N, N, N', N' tetramethylene diamine
Tris	tris-(hydroxymethyl)-methylamine
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside
AMPS	Ammonium persulfate
CIP	calf intestinal alkaline phosphatase
SAP	shrimp alkaline phosphatase
SW	salt water
TBE	tris-borate EDTA
OAA	oxaloacetate
Hb	*Halobacterium*
Ha	*Haloarcula*
Hf	*Haloferax*
HmMDH	*Haloarcula marismortui* malate dehydrogenase
DHFR	dihydrofolate reductase
THM-TF	formylmethanofuran: tetrahydromethanpterin formyltransferase
HmFd	*Haloarcula marismortui* ferredoxin

Acknowledgements

I would like to thank my project supervisors Prof. Michael Danson and Dr. David Hough for their guidance and supervision of this project over the last four years. Their role extended beyond the realms of an average supervisor, I think that you'd both give Claire Rayner a run for her money! I would also like to thank Dr Rupert Russell for his assistance with the creation of the molecular model and advice on the design of the site-directed mutants.

A big Thank You is owed to my family, Speed, Heinous, McT, Grandma and Grandpa for putting up with me and supporting me through my various PhD crisis's, I couldn't have done it without you. "Hey Dad," maybe now I'll stop "*FREELOADING*" and get a "*PROPER JOB*"! An equally big Thank you goes to Gareth, who gave up his leave to help me get my thesis together, I owe you big time. Thanks must also go to Louie and the girls for lending a shoulder to cry on through PhD black spots and providing the best little chef Chirton has ever known.

I would also like to thank members past and present of labs 1.33, and 1.28 with whom we share a dry area. Special mentions must go to Keithy, HC and Ursula for giving advice on laboratory techniques, Linus, Pete, Carlos, Krishty, Richie B, JJ, Squirtle, Queenie and Frankie Boy for providing constant on-line entertainment.

Last but not least, I would like to thank my chemistry A-level teacher Dr. Barnard, for without him I would not have passed my chemistry A-level all those years ago, and therefore would not be where I am today.

Financial support for this work was provided by the BBSRC.

For My Family
Speed, Heinous, McT, Grandma & Grandpa

Chapter 1

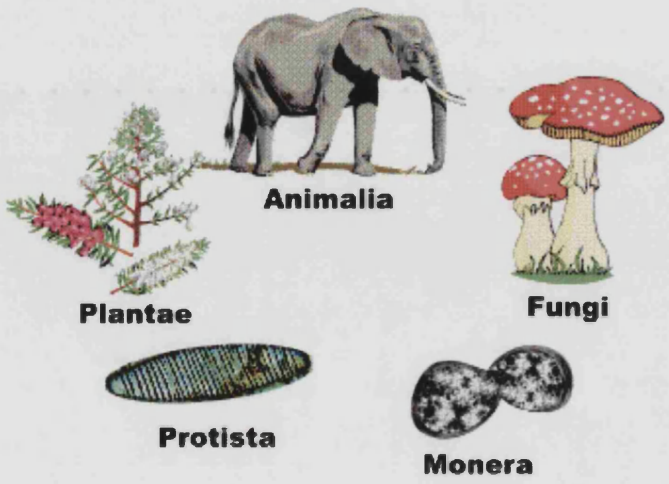
Introduction

1.1 Phylogenetic Analysis of the Archaea

1.1.1 Five Kingdom Paradigm

Traditionally, living organisms were separated into plants and animals, but it soon became apparent that many species of micro-organism could be classified as neither and a paradigm shift was required. The initial classification of all organisms was proposed by Ernst Haeckel in 1886, who formed a phylogenetic tree subdividing the living world into 3 kingdoms, Animalia, Plantae, and Protista (all microorganisms). The subsequent discovery of bacteria introduced a fourth branch and the subdivision of fungi culminated in the 5 kingdom hypothesis (Fig 1.1) Then came the realisation that all organisms could be defined cytologically on the basis of the possession of a nucleus. Thus, all organisms were divided into two basic types, prokaryotes organisms that lacked a cell nucleus and eukaryotes organisms that possessed a nucleus. From this was born the 5 kingdom classification where plants, animals, protista and fungi were classified under the term eukaryote, and bacteria under the term prokaryote.

Figure 1.1 **Five Kingdom Paradigm**



1.1.2 Phylogeny Prior to the Molecular Revolution

Phylogenetics is the historic evolutionary development of an organism or group of organisms. Traditionally this was compiled using morphology and the fossil record to develop evolutionary concepts. However, unlike that for macro-organisms, the microbial fossil record is sparse and difficult to interpret, and therefore phylogenetic trees are best constructed from comparisons of biochemical studies and sequences of informational macromolecules in living organisms.

1.1.3 rRNA as an Evolutionary Tool

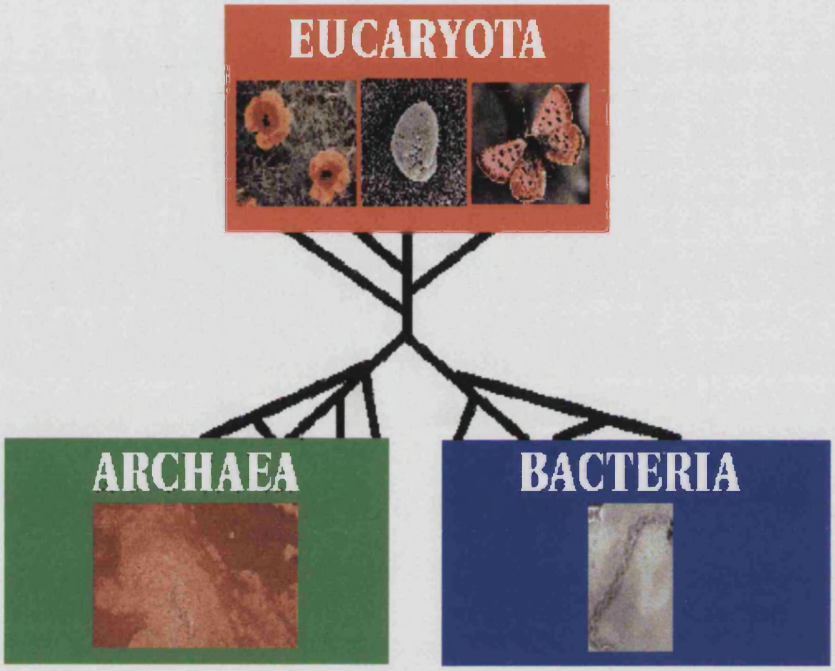
Until the advent of molecular biology, the traditional 5 kingdom division formed the cornerstone of biological classification. However, once DNA sequencing became a reality, the study of the 16S/18S rRNA of Prokaryotes and Eukaryotes has indicated that, from an evolutionary standpoint, the Prokaryote / Eukaryote dichotomy may not be a valid means of grouping organisms. Molecular phylogenetics was placed on an objective basis by Carl Woese when he conducted the first rigorous analysis of 16S and 18S rRNA of a wide range of organisms (Woese & Fox, 1977; Woese *et al.*, 1978).

1.1.4 Three Domain Paradigm

Studies of molecular evolution show that 2 evolutionary distinct groups of Prokaryotes exist, Eubacteria and Archaeabacteria which have subsequently been renamed as Bacteria and Archaea respectively. As the name implies "Archaea" are thought to reflect the earliest forms of life on earth and, consistent with this, many of the genera found within the domain Archaea are found in some of the earth's most extreme environments. Subject to the discovery of a third kingdom, Woese later proposed that a new taxon be introduced above the level of a kingdom called a "Domain", with all life comprising three domains: the Bacteria, the Archaea and the Eucarya, each containing at least two kingdoms (Woese *et al.*, 1990).

The three domain theory, originally suggested by Woese in 1977, has since been supported by further phylogenetic studies based on 5S rRNA (Hori & Osawa, 1982), 16S and 18S rRNAs (Gouy & Li, 1989) and RNA polymerasees (Zillig *et al.*, 1989).

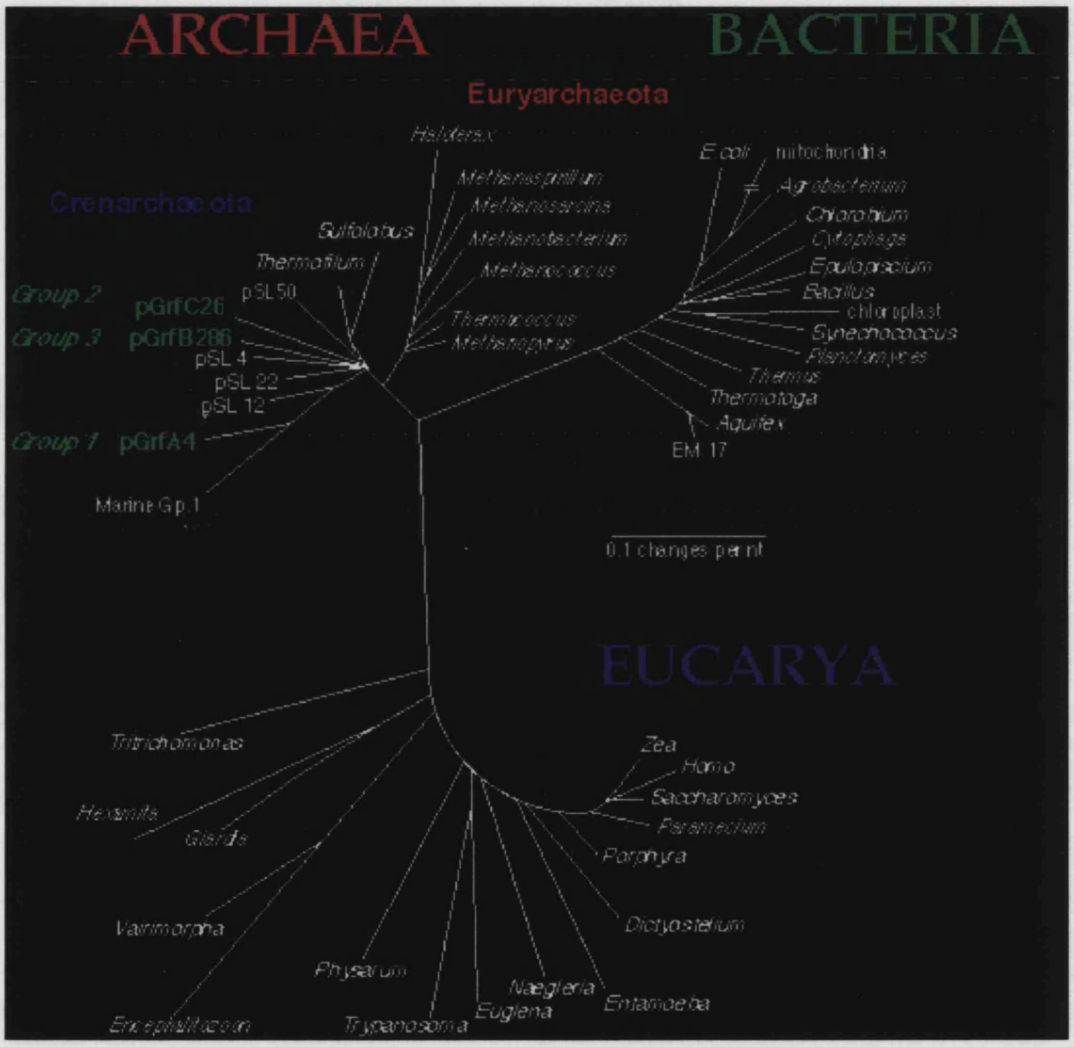
Figure 1.2 THREE DOMAINS OF LIFE



1.1.5 Euryarchaeota, Crenarchaeota and Korarchaeota

The archaeal domain comprises three kingdoms, Euryarchaeaota, Crenarchaeota and Korarchaeota. Euryarchaeota are a phenotypically-diverse group containing the methanogens, the extreme halophiles, sulphate-reducers and some thermophiles. The kingdom Crenarchaeota is comprised mainly of hyperthermophiles, and has the distinction of including microbial species with the highest known growth temperatures of any organism (Barns *et al.*, 1994, 1996). However, rRNA sequence based analyses indicate that Crenarchaeota may also be distributed in low-temperature environments such as ocean waters and terrestrial sediments and soils (Bintrim, 1997, DeLong, 1994, Furham, 1992, and Hersberger, 1996). The Korarchaeota are a recently discovered addition to the archaeal domain and the group currently includes anonymous organisms identified by environmental sample sequences isolated from hot springs (Barns *et al*, 1996). The position of the Euryarchaeota and Crenarchaeota within an unrooted form of the 16S rRNA phylogenetic tree can be seen in Figure 1.3.

Figure 1.3 The 16S rRNA Universal Tree in an Unrooted Form



Unrooted form of the phylogenetic tree produced using the distance matrix alorithm. Based on 16S rRNA sequences collected by Woese et al. (1978).

A summary of the major milestones that have shaped our current understanding of biological classification can be seen in Table 1.1.

Table 1.1
Chronological Compilation of Biological Classification

<table>
<tr><th>two kingdoms</th><th>Five kingdoms</th><th colspan="2">three domains and who knows how many kingdoms</th></tr>
<tr><td>Animalia</td><td>Animalia</td><td rowspan="11">Eukarya</td><td>Animalia</td></tr>
<tr><td rowspan="2">Plantae</td><td>Fungi</td><td>Fungi</td></tr>
<tr><td>Plantae</td><td>Plantae</td></tr>
<tr><td rowspan="8">Either Protozoa (=Animal) or algae (=Plant)</td><td rowspan="8">Protocista</td><td>Alveolata</td></tr>
<tr><td>Stramenopiles</td></tr>
<tr><td>Etc...</td></tr>
<tr><td>Sporozoa</td></tr>
<tr><td>Mycetozoa</td></tr>
<tr><td>Euglenozoa</td></tr>
<tr><td>Etc...</td></tr>
<tr><td>Archezoa</td></tr>
<tr><td rowspan="2">Plant (bacteria and blue-green algae)</td><td rowspan="2">Monera</td><td>Eubacteria</td><td>(kingdoms not specified)</td></tr>
<tr><td>Archaea</td><td>Euryarchaeota
Korarchaeota
Crenarchaeota</td></tr>
</table>

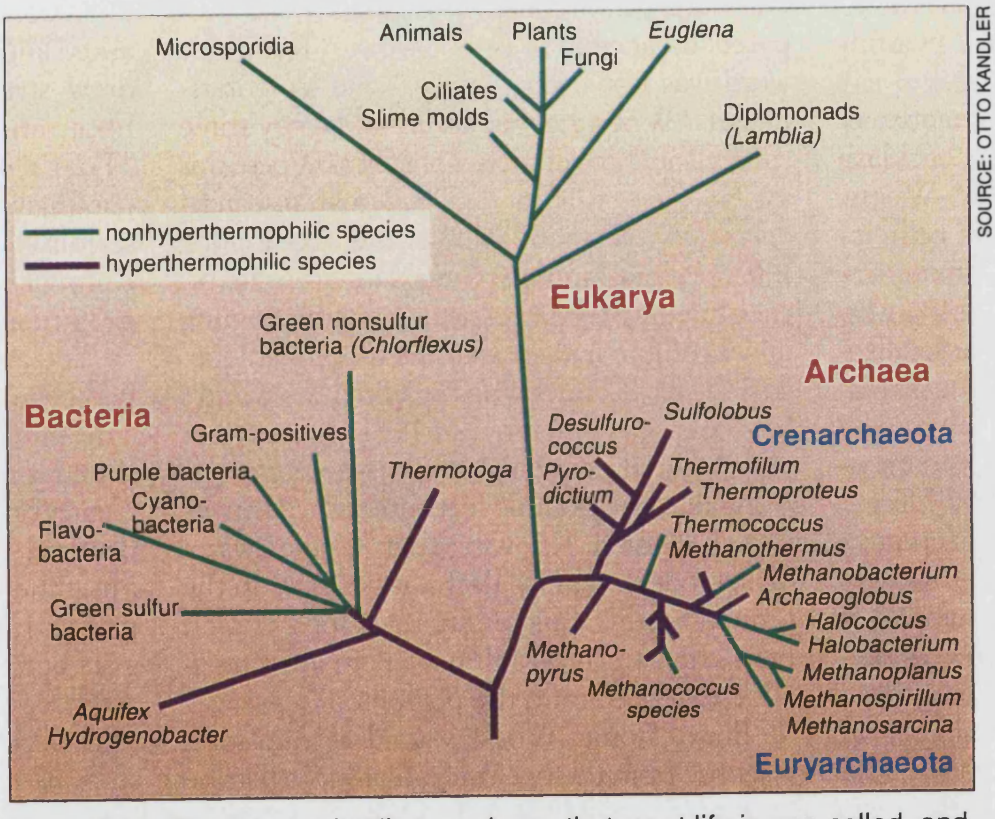
1.1.6 Rooting of the Universal Tree

A phylogenetic tree derived from a comparison of a single molecular species from a range of organisms is generally unrooted, because the universal ancestor can only be supposed and not confirmed. A sequence-based tree relating a restricted group of organisms can be rooted by establishing at which point it is closer to an outgroup, but for a tree relating all organisms there can be no such outgroup (Doolittle & Brown, 1994).

Two research teams independently discovered a method to root the universal tree via ancient duplicated gene products. One group used elongation factors EF-Tu and EF-G and the α and β subunits of F1-ATPase, F1-α and F1-β (Iwabe *et al.*, 1989); the second group used H^+ ATPase submits (Gogarten *et al.*, 1989). The phylogenetic tree derived from 16S rRNA was rooted (Woese *et al.*, 1990; Wheelis *et al.*, 1992), using the method of Iwabe and Gogarten named the Dayhoff strategy. By inferring a composite phylogenetic tree from genes that show an extensive sequence homology across members of the three domains, the root can be positioned where the two genes are thought to have diverged.

Since the rooting of the 16S rRNA tree, others have tried the same method on different gene families but found that it was impossible to root the tree conclusively (Benachenhou-Lahfa *et al.*, 1993; Forterre *et al.*, 1993). In response to criticism of the work by Gogarten and Iwabe, a root for the universal tree was derived using aminoacyl-tRNA synthetase genes (Brown & Doolittle, 1995). These genes form part of an extensive multigene family whose duplication is believed to have preceeded the divergence of the three primary domains. Their findings confirmed the position of the root being placed between archaebacterial/eukaryotes and eubacteria, thus, supporting the Archaeabacterial Tree based on 16S rRNA (Fig 1.4) and the protein topologies by Iwabe *et al.* (1989), and Gogarten *et al.*(1989).

Figure 1.4 The 16S rRNA Universal Tree in Rooted Form



Tree of life. The Woese family tree shows that most life is one-celled, and that the oldest cells were hyperthermophiles.

Tree reproduced from "*Microbiology's scarred revolutionary*", Science, 1997

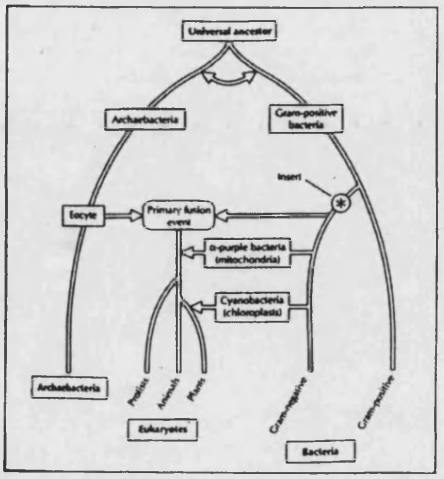
1.1.7 Alternatives to Woeses Universal Tree

The position of the Archaea within the universal tree still remains a controversial topic of debate and the universal phylogenetic tree of Woese (1978), has been challenged by other members of the evolutionary community.

A classification based solely on differences in rRNA sequences ignores the apparently enormous evolutionary step from prokaryotes to the eukaryotes (Mayer, 1990). The assumption that molecules of a single member of a species yield information typical of that species was questioned (Margulis & Guerrero, 1991; Margulis, 1996). Concern is shown when the chosen sequence varies between different tissues of an organism, or at different stages in its life cycle. Evolutionary studies of glutamate dehydrogenase presented unusual branching patterns of the Archaea (Labedan *et al.*, 1993) and subsequent studies on glutamine synthetase (Brown *et al.*, 1994), and heat shock protein 70 kDa (Gupta & Singh, 1994) also refute Woese's universal tree.

In an attempt to explain the contradictory phylogenies based on different gene sequences Gupta & Singh (1994) proposed a chimera model for the origin of the eukaryotic cell involving a fusion event between a Gram-negative bacterium and a Crenarchaebacterium which they term an eocyte; a group of extremely thermophilic, sufur-metabolizing, anucleate cells (Lake *et al.*, 1984).

Figure 1.5 The Chimera Model



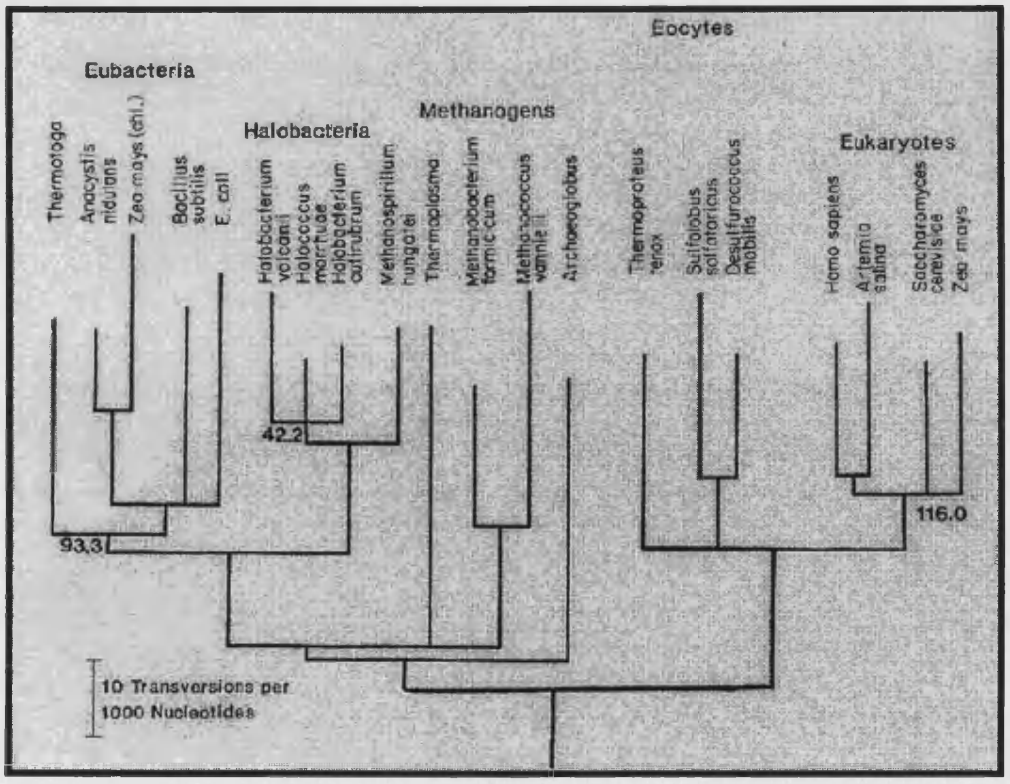
Reproduced from Gupta & Singh (1994)

1.1.8 James Lake's Eocyte Tree

The Eocyte tree was produced by James Lake (Lake, 1988,1989) as an alternative to Woese's universal tree. It was produced by analysis of 16S rRNA sequences using a method called Evolutionary Parsimony, an algorithm invented by Lake himself. The method is claimed to be insensitive to unequal rates of evolution between different organisms and thus perfect for the analysis of highly diverged sequences. He argues that all phylogenetic trees produced without this method are artifacts due to the unequal rate phenomenon. The tree obtained by this algorithm suggests two superkingdoms, with the methanogens and halophiles grouping with the eubacteria and the sulfur-dependent thermoacidophiles grouping with the eukaryotes. In this proposal, the Archaea are polyphyletic.

This method has subsequently been criticised (Olsen & Woese, 1993) on the basis of the alignment used and other methods employed give alternative results. Evolutionary parsimony used on 23S rRNA gives a completely different grouping to Lake (Gouy & Li, 1989).

Figure 1.6 Lake's Eocyte Tree



Reproduced from Lake (1988)

1.1.9 Are Archaea the Earliest Forms of Life?

The Search for the Origin of Life Continues

The search for the origin of life continues and the manifestation of the universal ancestor of all extant life on earth still remains unknown. It is however accepted by the majority of the phylogenetic community that life on Earth can be divided into three domains, with the position of the root of the universal tree of life still being fraught with ambiguity.

The debate over the early evolution of life still remains complex and diverse. Implications of early evolution of life can be obtained essentially from the physical (atmospheric, geological) and biological (fossil and molecular) records. Theories toward a high temperature early evolution of life in environments analogous to present day geothermal and submarine hydrothermal vents are focussing more attention. If life did originate in such an environment (Baross & Hoffmann, 1985; Pace, 1991; Wächtershauser, 1988,1989) , the prevailing niche should be very similar to that of early earth, namely, hot and reducing (Holland, 1984; Pace, 1991) and possibly the first metabolic pathways required these conditions. In the reducing atmosphere of geothermal systems, chemolithoautotrophy is a prevalent process among the thermophilic microorganisms, and therefore a chemoautotrophic origin for life (Wächteshauser, 1988, 1990) is tenable theory. Others would argue that the concentration of salts and organic compounds present would make the earliest life forms halophilic in nature (Dundas, 1977; Stoeckhenius, 1978). Recent unveiling of nano-organism (nanobe) fossils from Australian sandstones (Unwins *et al.*, 1998) presents evidence that could indicate survival from a meteoritic bombardment, and they therefore could be the "universal ancestor" of all extant life on Earth. The debate whether life on Earth first evolved in hydrothermal environments (Forterre, 1996; Gogarten-Boekels *et al.*, 1995) or merely survived meteoritic bombardment, still remains an open question.

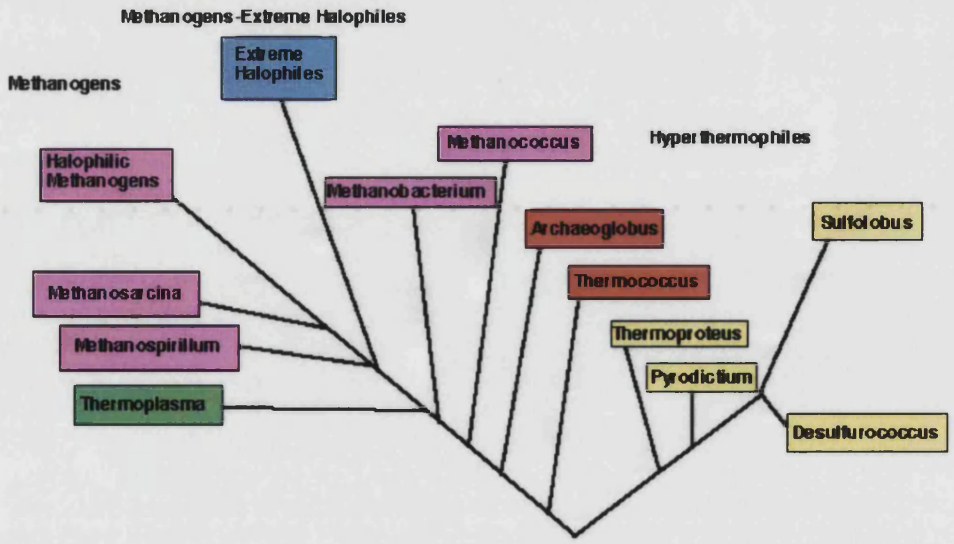
1.2 Extremophiles, Microorganisms at the Frontiers of Life.*

1.2.1 Archaeal Phenotypes

Environments that from the human vantage are considered to be extreme are colonized by special microorganisms which are adapted to these ecological niches; such organisms are termed extremophiles. "Extremophiles" encompass certain species of bacteria and the majority of the domain Archaea.

Originally there was thought to be three main phenotypes of Archaea, namely thermophilic, halophilic and methanogenic, (Fig 1.7) (reviewed by Woese & Olsen, 1986, Fewson, 1986, Woese, 1987 and Danson *et al.*, 1992). However, a fourth phenotype was added when a psychrophilic Archaeon was discovered (Preston *et al.*, 1996) and a fifth upon the recent discovery of mesophilic Archaea.

Figure 1.7 Tree of Archaeal Phenotypes



Key: Methanogens – pink Halophiles – blue Thermophiles-green & orange sulfur-dependant thermophiles-yellow

* Photos taken by NASA. L – R Deep Ocean vent community, Algal mats on an Antarctic lake bottom, endolithic algae from Antarctica, hot springs in Yellowstone National Park.

1.2.2 Thermophiles*

Thermophiles are amongst the best studied of the Archaea since first being discovered in Yellowstone Park by Brock in the 1960's. They are found in habitats of temperatures between 55 – 110 °C, and such environments include fumaroles and geysers, shallow submarine and hydrothermal abyssal hot-vent systems (black smokers), and pools located in the vicinity of active volcanos.

Sulfurous acid springs
White Island, (New Zealand)

Juan de Fuca smoker, Pacific Ocean , Mexico

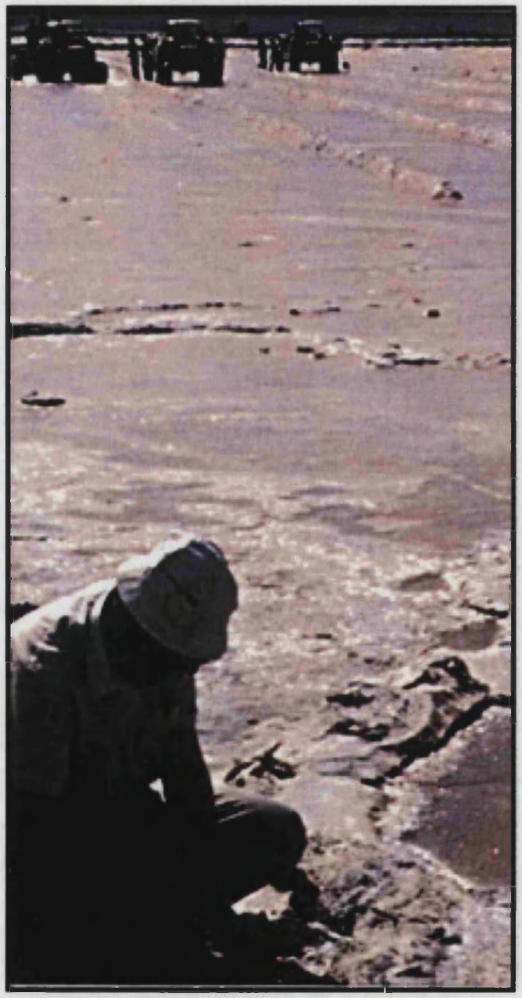
Some thermophiles possess the ability to metabolize sulfur in some way. Due to their inclination to inhabit solfatara and sulfur pyrite areas some terrestrial thermophiles posses the ability to grow in very acidic environments, certain acidophiles have even been isolated at pHs of below 0. They are also found at the opposite end of the spectrum in extreme alkaline environments including alkaline hot spas, carbonate springs, alkaline soils and soda lakes.

Low pH drainage stream from a copper mine (Sardinia).

* Illustrations from Biology of extremophiles brochure, Biotech programme of the European Union.

1.2.3 Methanogens

The methanogens are obligate anaerobes converting CO_2 to methane, (Jones *et al.*, 1987), and they are found in habitats such as bogs, pond sediments, sewage lagoons and intestinal tracts of animals. Methanogens inhabiting a cow's intestine are estimated to release in the vicinity of 5L of methane a day.



Extreme alkaline lake saturated with NaCl. Lake Magadi (Kenya).

1.2.4 Halophiles

As the name suggests, halophiles require high salt concentrations, and have thus evolved mechanisms to withstand such environments. Their intracellular concentration of salts are isotonic with that of their external environment. The predominant cations that halophiles adapt to and utilize are Na^+, K^+, Mg^{2+}, and Ca^{2+} (Dennis, 1997).

1.2.5 Psychrophiles

The habitats of psychrophilic organisms include the cold polar seas Antartic soils and alpine glaciers, as well as deep-sea sediments which are not only permanently cold but are also at high pressure.

1.2.6 Mesophilic Archaea

Molecular genetic methods which enable detection of microorganisms without the need to cultivate or isolate them, have shown that mesophilic Archaea are ubiquitously distributed inhabiting environments with a broad spectrum of temperatures. Their presence has been shown in the upper layers of the oceans (DeLong, 1992, Fuhrman *et al.*, 1992), in freshwater sediments (Hershberger *et al.*, 1996, Schleper *et al.*, 1997, MacGregor *et al.*, 1997), in a meromictic lake (Ovreås *et al.*, 1997) and in different soil types (Ueda *et al.*, 1995, Bintrim *et al.*, 1997, Kudo *et al.*, 1997, Borneman & Triplett, 1997), including boreal forest soil (Jurgens *et al.*, 1997, Jurgens & Saano, 1999).

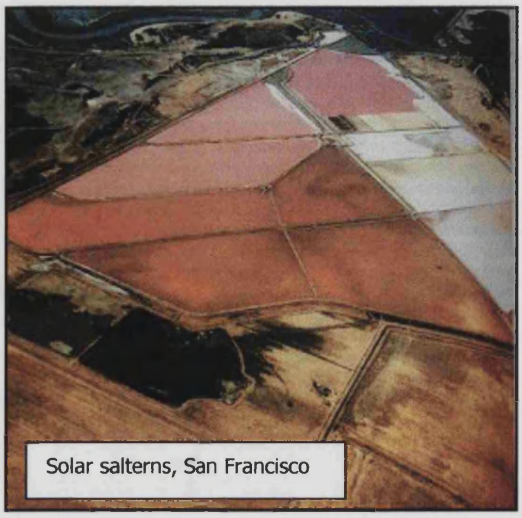
Extremophilic microorganisms often live in biotopes combining selective stress factors, for instance high temperature and acidic conditions. It is not unreasonable to speculate that, wherever there is an extreme environment, there are very likely to be extremophiles. Continuing exploration of different extreme niches is persistently revealing additional species of Archaea, confirming that this Domain is indeed diverse and studies into this area need to be constantly updated (*Pace et al.*, 1996).

1.3 The Halophilic Archaea

Part of Lake Magadi, Kenya. Environments are highly saturated with salt and crystallisation occurs on the surface

The extreme halophiles are a phenotype of the domain Archaea that require high concentrations of salt. The term extreme is used to emphasize that these organisms are not just halophilic, but that their requirement for salt is extreme, in some cases near saturation. Halophiles grow optimally at salt concentrations of up to 4.5M (Robb *et al.*,1995), with the majority requiring between 3-4M. Virtually all extreme halophiles can grow at 5.5M, which is the limit of saturation for NaCl, although some species only grow very slowly at this concentration.

Environments where halophiles are likely to be found include marine salterns, solar salt evaporation ponds, hypersaline lakes such as the Great Salt Lake and the Dead Sea. Halophiles have also been isolated from artificial saline habitats such as highly salted fish, certain sausages and salted meat or pork (Madigan *et al.*,1997).



Solar salterns, San Francisco

1.3.1 Taxonomy and Physiology of the Halophilic Archaea

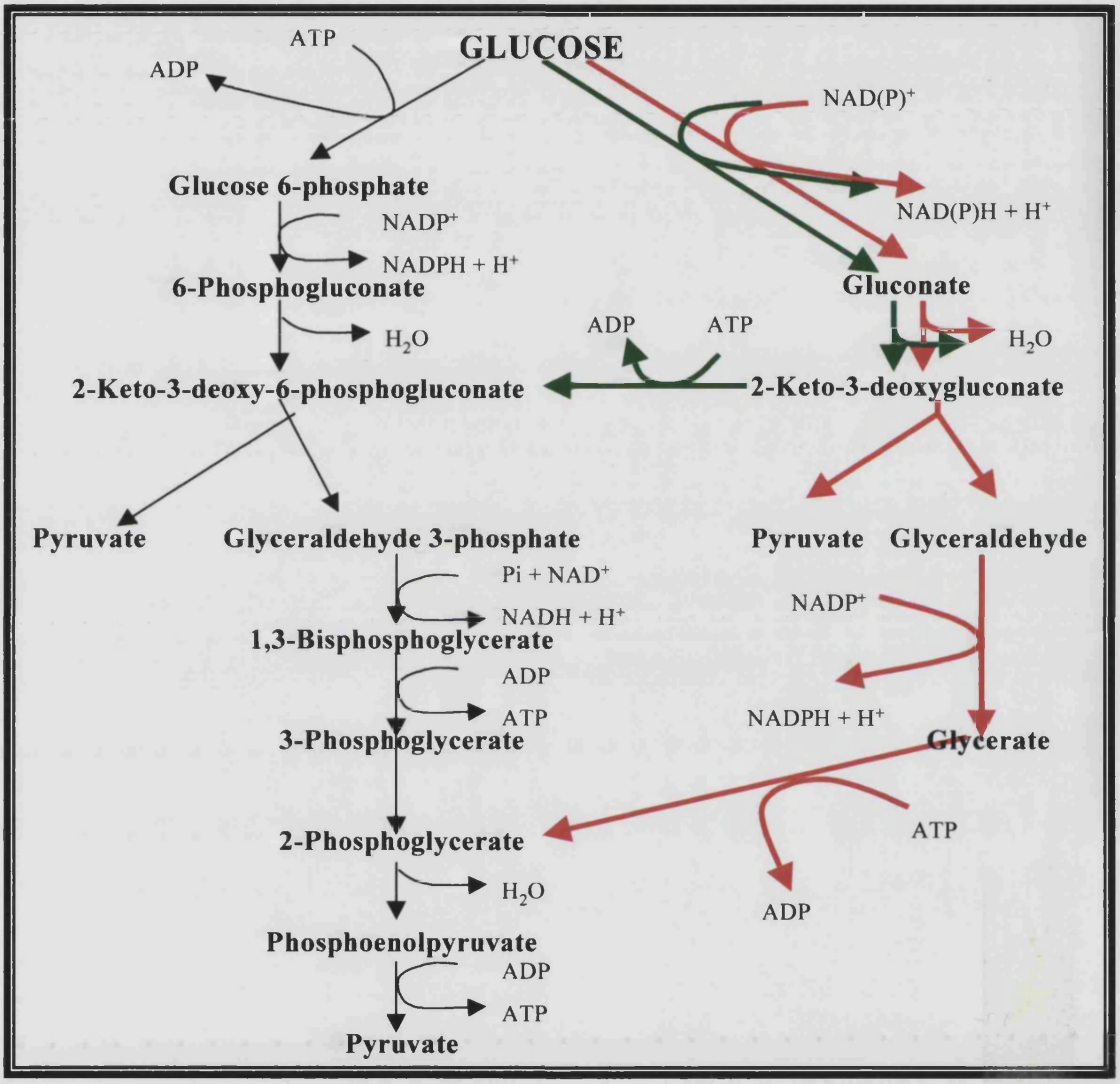
Ribosomal RNA/DNA hybridisation studies have divided the halophilic archaea into six genera, *Halobacterium*, *Halococcos*, *Haloferax* and *Haloarcula* growing at neutral pH, and *Natronobacterium* and *Natronococcus* which require alkaline conditions (Grant & Larsen, 1989; Rodriguez-Valera, 1988).

Like other forms of Archaea, halophilic Archaea lack peptidoglycan in their cell envelope S-layer consequently leading to a less rigid cell wall (Kushner *et al.*,1964, Brown & Cho, 1970). They are insensitive to most antibiotics concerned with cell wall and protein synthesis (Woese *et al.*, 1978, Bonelo *et al.*, 1984) with the exceptions of mevinolin, which inhibits 3-hydroxy-3-methylglutaryl CoA (HMG-CoA) reductase, present in the biosynthesis of isoprenoids and cholesterol, and novobiocin which is a DNA gyrase inhibitor.

Halophilic Archaea contain plasmids that are amongst the largest naturally-occurring plasmids known and which can contain up to 30% of the total cellular DNA. Halophilic genomes have a high GC content , which can be as high as 70%. In addition to the large amount of non-chromosomal DNA, some genera of halophilic Archaea contain a considerable amount of highly repetitive DNA, the function of which is unknown.

Halophilic Archaea are obligate aerobes which are capable of growing on complex media. They posses some distinctive biochemical pathways involved with central metabolism (Danson, 1993). Certain halophiles have the ability to catabolise glucose and galactose and it appears that they do so via a modified Entner-Doudoroff pathway (Tomlinson *et al.*, 1974). Glucose is converted to 2-keto-3-deoxygluconate, which is phosporylated to form 2-keto-3-deoxy-6-phosphogluconate and this in turn is cleaved to yield pyruvate and glyceraldehyde-3-phosphate (Fig 1.8)

Figure 1.8 Modified Entner-Doudoroff Pathway of Halophiles



Green lines = modified pathway of halophiles, Red lines = non-phosphorylated pathway of *Sulfolobus sulfataricus* and *Thermoplasma acidophilum*. Black lines = classical Entner-Doudoroff pathway of bacteria.
Reproduced from Danson (1993).

Certain species of extreme halophiles can utilise solar energy by a light-mediated synthesis of ATP, through their purple-membrane pigment bacteriorhodopsin. Other retinol proteins have also been identified including a light-driven chloride pump, halorhodopsin (Oesterhelt & Titter, 1989), and 2 sensory rhodopsins mediating the phototrophic response (Spudich, 1993). In addition to retinol proteins, halophiles also contain carotenoid pigments which cause the red to yellowish colours characteristic of halophilic blooms.

Figure 1.9 Characteristic Halophilic Blooms

Intense bloom of the halophilic Archaeon, *Halobacterium,* at a salt works San Quintin, Mexico. The intense red colour is due to the presence of retinoid pigment bacteriorhodopsin.

1.3.2 Phylogeny of Halophiles

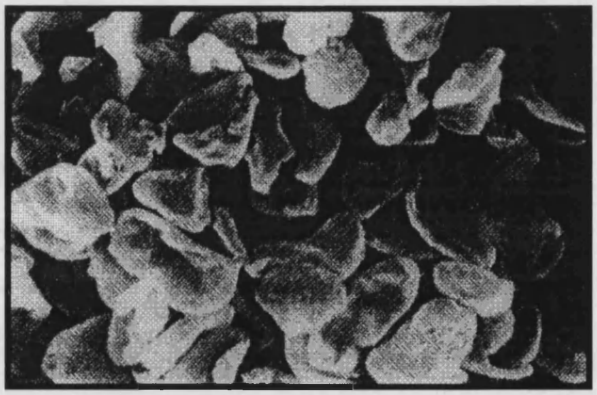
Halophiles are phylogenetically placed in the archaeal sub-division euryarchaeota. All algorithms involved in elucidating phylogenetic trees place the halophiles closest in evolutionary divergence to the methanogens, which is an interesting observation in that it has been proposed that the earliest living cells to evolve may have been halophilic in nature, resulting from the concentration of both salts and organic compounds in an evaporitic environment giving rise to the "primordial soup" (Dundas, 1977, Stoeckenius, 1978). If the prebiotic atmosphere was reducing as hypothesised, then the earliest halophiles would have been expected to have been obligate anaerobes with similar biochemical pathways to modern day methanogens.

1.4 Haloferax volcanii

1.4.1 Evidence of Life in the Dead Sea

Haloferax volcanii (*Hf.volcanii*) was first isolated in 1940 as evidence of life in the Dead Sea by the Israeli microbiologist Ben Volcanii. It has since been fully characterised (Mullakkhanbhai & Larsen, 1975; Torreblanca *et al.*, 1986). Cells are extremely pleomorphic, frequently disk or cup-shaped even under optimal conditions. It requires at least 1M NaCl for growth, with cells lysing upon contact with water. Optimum growth conditions are salt concentrations (NaCl/KCl) between 1.5 and 2.5M at 30-50°C. These salt conditions categorise *Haloferax* as a moderate-extreme halophile. *Haloferax* is capable of growing on a simple, defined minimal medium, allowing the isolation of auxotrophic strains (Mevarech & Werezberger, 1985). The organism has a high requirement for Mg^{2+}, with a minimal concentration of 0.02M (Torreblanca *et al.*,1986), an optimum concentration of 0.2M, and a tolerance of up to 1.5M. Its requirement and tolerance to Mg^{2+} is higher than many other halophiles.

Figure 1.10 Haloferax volcanii strain DS2



Photographed by the Laboratory of Clinical Electron Microscopy, *University of Bergen.*
Reproduced from Bergey's manual.

Chemoorganotrophic growth occurs on sugars as the sole source of carbon (Larsen, 1984). Acid is produced from glucose and, by implication, from a variety of sugars (Rodriguez-Valera *et al.*, 1983).

1.4.2 Molecular Biology of Haloferax volcanii

The DNA is composed of a major and minor component, with the presence of plasmid DNA as well as chromosomal DNA (Pfeifer *et al.*, 1981; Gutirrez *et al.*, 1986). The genome of *Hf.volcanii* consists of a circular 2.9 Mb chromosome and four plasmids of 690, 442, 89 and 6.4 kb. Detailed physical mapping of the *Hf.volcanii* genome has yielded a set of overlapping cosmid clones that cover 96% of the total sequence (Charlebois *et al.*, 1990, 1991). The mol% G + C of the major DNA component is 63.4 (T_m) and 66.5 (Bd), and that of the minor component, 55.3 (Bd). The *Haloferax* genome contains fewer mobile insertion sequences than those of other halophilic Archaea making it genetically relatively stable.

Transformation techniques have been developed for *Hf.volcanii* (Charlebois *et al.*, 1987) and a number of shuttle vectors are now available that will replicate in both *E.coli* and *Hf.volcanii* (Lam & Doolittle, 1989; Holmes *et al.*, 1991; Holmes *et al.*, 1994; Jolley *et al.*, 1997; Connaris *et al.*, 1999). These developments have established the organism as perhaps the species of choice for genetic studies of halophilic Archaea.

1.5 Halophilic Enzymes

1.5.1 The Internal Salt Environment of Halophilic Archaea

Organisms that grow in hypersaline environments have developed various mechanisms to overcome the extracellular osmotic pressure. It is now known that to establish an osmotic balance with the environment, the Archaea maintain internal levels of salt, mainly KCl, that are isotonic with their environment (Kusher, 1988); consequently, the entire biochemical machinery of halophilic archaebacteria must be adapted to function at high salt concentrations (Eisenberg *et al.*, 1992). This method of osmotic adaptability differs from the mechanism found in halophilic eubacteria and eukaryotes, which synthesize large quantities of small organic osmoprotectants (Borowitzka, 1974). These low molecular weight organic solutes, otherwise known as compatible solutes, act to lower the osmotic potential, (ψ) of the cell, thus attracting water back into it. Therefore, enzymes from the halophilic archaea must function in near-saturating concentrations of KCl. However, not only can their enzymes withstand high concentrations of salt, but in many cases they actually require such levels both for stability and enzymatic activity.

The early observations made by Reistad (1970) and Lanyi (1974) regarding the excess nature of acidic amino acid residues and lowered content of hydrophobic residues present in protein from extreme halophiles provided the infrastructure for the current theories on protein halophilicity. However, in order to formulate hypotheses and theories on the molecular basis of protein halophilicity, it is essential to understand the effects and interactions that take place when a non-halophilic protein comes into contact with excess concentrations of salt.

1.5.2 Effect of Salt on Non-halophilic Protein Structures

Surface contact between protein and water constitutes an interface, and thus a surface tension exists due to the cohesive hydrogen-bonding nature of the solvent; this in turn leads to the characteristic globular shape of proteins through their attempting to attain a minimal surface area and, ultimately, a minimal surface energy. Changes to the solvent that affect the surface tension will therefore also affect protein conformation by altering the cohesive force of water and hence the surface tension.

Most inorganic salts increase the surface tension of water. The surface area of a denatured protein is greater than that of the native protein and therefore the increase in surface tension of the water promotes the formation of a compact globular protein. However, as the concentration of salt is increased, so the surface tension of the water increases and consequently the protein eventually aggregates to reduce the solvent-accessible surface area, resulting in salt precipitation (Timasheff,1992).

The problem faced by halophilic proteins is one of remaining soluble. To do this it must attract the salt and water back to its surface and so prevent aggregation.

1.5.3 Current Hypotheses on Halophilicity

The present theories on halophilic protein stability have essentially been formed around 4 halophilic structures of halophilic enzymes [malate dehydrogenase (HmMDH) from *Haloarcula marismortui* (Dym *et al.*, 1995), dihydrofolate reductase (DHFR) from *Haloferax volcanii* (Pieper *et al.*,1997), tetrahydromethanopterin formyltransferase (THM-TF) from *Methanopyrus kandlerin* (Ermler *et al.*,1997) and ferredoxin (HmFd) from *Haloarcula marismortui* (Frolow *et al.*, 1996)] and a handful of sequences and molecular models (Danson

& Hough, 1997; Rao & Argos, 1981; Eisenberg *et al.*,1992). For images of the structures see Figure 1.11.

The structures all have the characteristic high acidic amino acid content initially observed by Lanyi (1974), now put spatially into context with the generation of 3-dimensional structures. In all crystal structures and molecular models studied, the majority of acidic residues are seen to be placed on the surface of the molecule. This phenomenon is presumably allowing the proteins to counteract the aforementioned salting out reaction in conditions of extreme salinity, by attracting the water molecules back to the surface of the protein in the form of hydrated salt ions. This is thought to be one of the main adaptions for stability for all known halophilic structures, although Bohm and Jaenicke (1994) have stressed that instability at low salt concentrations (through charge-charge repulsion) might be an alternative way of viewing protein halophilicity. It has also been proposed that hydrophobic interactions play an important role in the ability of these proteins to cope with the salt stress in a hypersaline environment (Eisenberg & Watchel, 1987; Jaenicke, 1991; Eisenberg *et al.*, 1992).

Halophilic enzymes isolated at the time this table was constructed (1997) are seen in Table 1.2.

Table 1.2 Enzymes Isolated from Extreme Halophiles

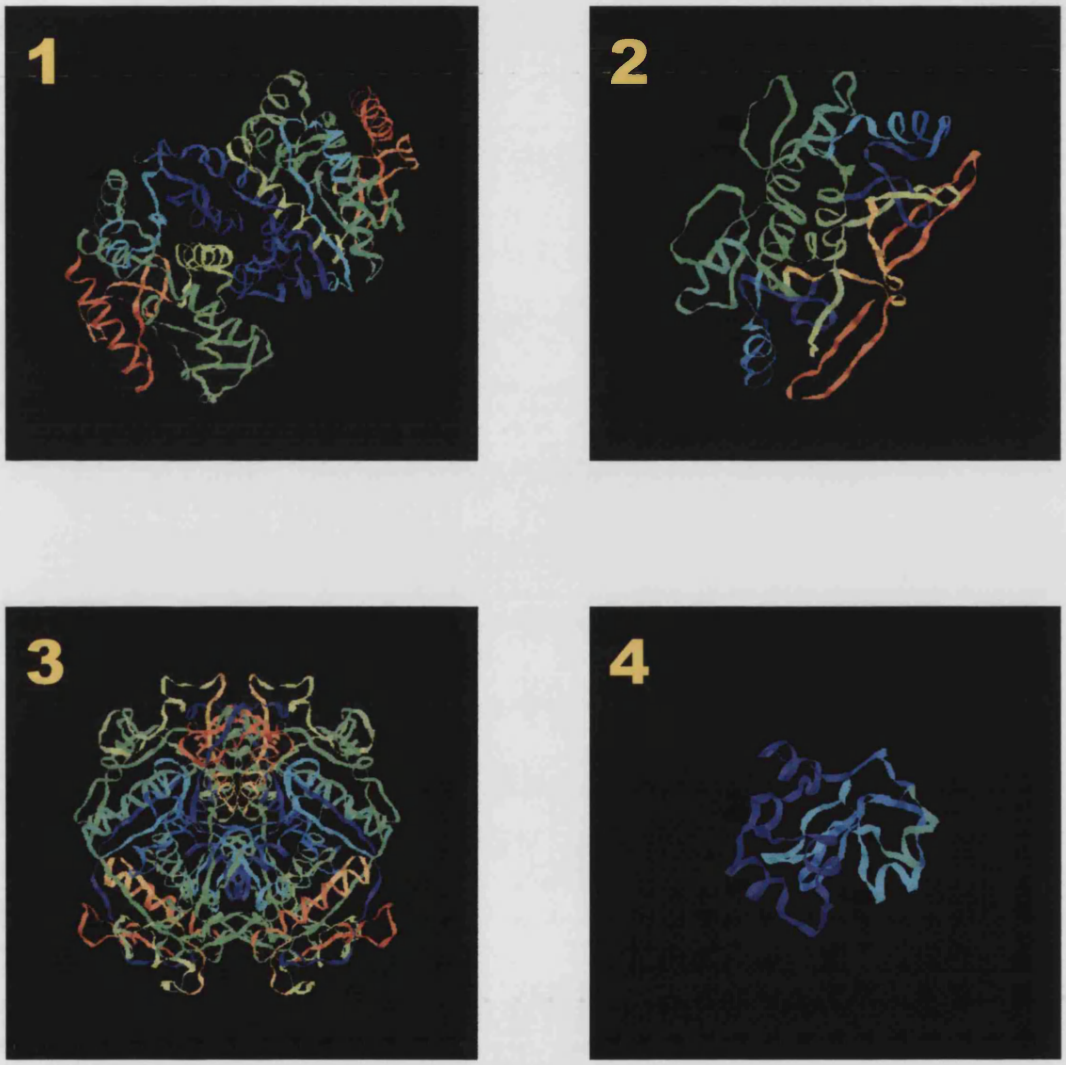
	ENZYME	ORGANISM	REFERENCE
OXIDOREDUCTASES	Alanine dehydrogenase	*Hb. salinarium*	Keradjopoulos & Holldorf, 1979
	Catalase	*Hb. cutirubrum*	Lanyi & Stevenson, 1969
	Dihydrolipoamide dehydrogenase	*Hb. halobium*	Danson *et al.*, 1986
	Malate dehydrogenase	*Ha. marismortui* *Hb. salinarium*	Mevarech *et al.*, 1977 Holmes *et al.*, 1965
	Peroxidase	*Hb. halobium L-33*	Fukumori *et al.*, 1985
	Superoxide dismutase	*Hb. cutirubrum* *Hb, halobium*	May & Denis, 1987 Salin & Oesterhelt, 1988
TRANSFERASES	Aspartate carbamoyltransferase	*Hb. cutirubrum*	Norberg *et al.*, 1973
	Pyruvate kinase	*Hb. cutirubrum*	De Medicis *et al.*, 1982
	RNA polymerase	*Hb. cutirubrum* *Hb. halobium*	Louis & Fitt, 1972 Zillig *et al.*, 1978
HYDROLASES	Amylase	*Hb. halobium*	Good & Hartmann, 1970
	Amyloglucosidase	*Hb. sodomense*	Oren, 1983
	ATP synthase	*Hb. halobium* *Hb. saccharovorum* *Hf. mediterranei* *Hf. volcanii*	Nanba & Mukohata, 1987 Kristjansson & Hochstein 1985 Dane *et al.*, 1992 Dane *et al.*, 1992
	Phosphatase	*Hb. halobium*	Bonet *et al.*, 1992
	Protease	*Hb. halobium*	Ryu *et al.*, 1994
LYASES	Aldolase	*Ha. vallismortis*	Krishnan & Altekar, 1993
	Ribulase 1,5-bisphosphate carboxylase	*Ectothiorhodospira halophila*	Tabita & McFadden, 1976

Reproduced from conference talk: How are Genes Expressed in Extreme Halophiles?, Richard F. Shand, *Dept Biological Sciences, Northern Arizona University.*

Key:

Ha. = Haloarcula
Hb. = Halobacterium
Hf. = Haloferax

Figure 1.11 Halophilic Enzyme Crystal Structures



1 Malate Dehydrogenase (HmMDH) from *Haloarcula marismortui* (Dym, *et al.*, 1995)

2 Dihydrofolate Reductase (DHFR) from *Haloferax volcanii* (Pieper *et al.*, 1997)

3 Formylmethanofuran: tetrahydromethanopterin formyltransferase (THM-TF) from *Methanopyrus kandleri* (Ermler *et al.*, 1997)

4 Ferredoxin (HmFd) from *Haloarcula marismortui* (Frolow *et al.*, 1996)

All files were imported from the Ebi PDB and reproduced using RasMol

1.6 Citrate Synthase

1.6.1 The Role Played in Central Metabolism

Many of the central metabolic pathways have been found in all three domains, and it is believed that these pathways were established before the separation of the three evolutionary lineages. Already central metabolic comparative enzymology has played an important role in understanding cellular evolution and the relationship between protein structure and function (Danson, 1988; 1993).

The citric acid cycle is the site of the terminal reactions of aerobic catabolism in Eukarya, Archaea and Bacteria, acting as a source of reducing equivalents for ATP synthesis (Fig 1.13). The cycle also provides a pool of biosynthetic precursors for anabolic processes. Citrate synthase acts as a major control point in the citric acid cycle, facilitating the entry of carbon into the cycle by catalysing the condensation of oxaloacetate (OAA) and acetyl coenzyme A to form citrate and coenzyme A (Fig 1.12).

Figure 1.12 Citrate Synthase Reaction

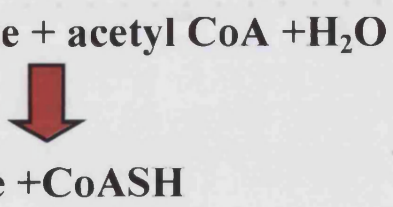
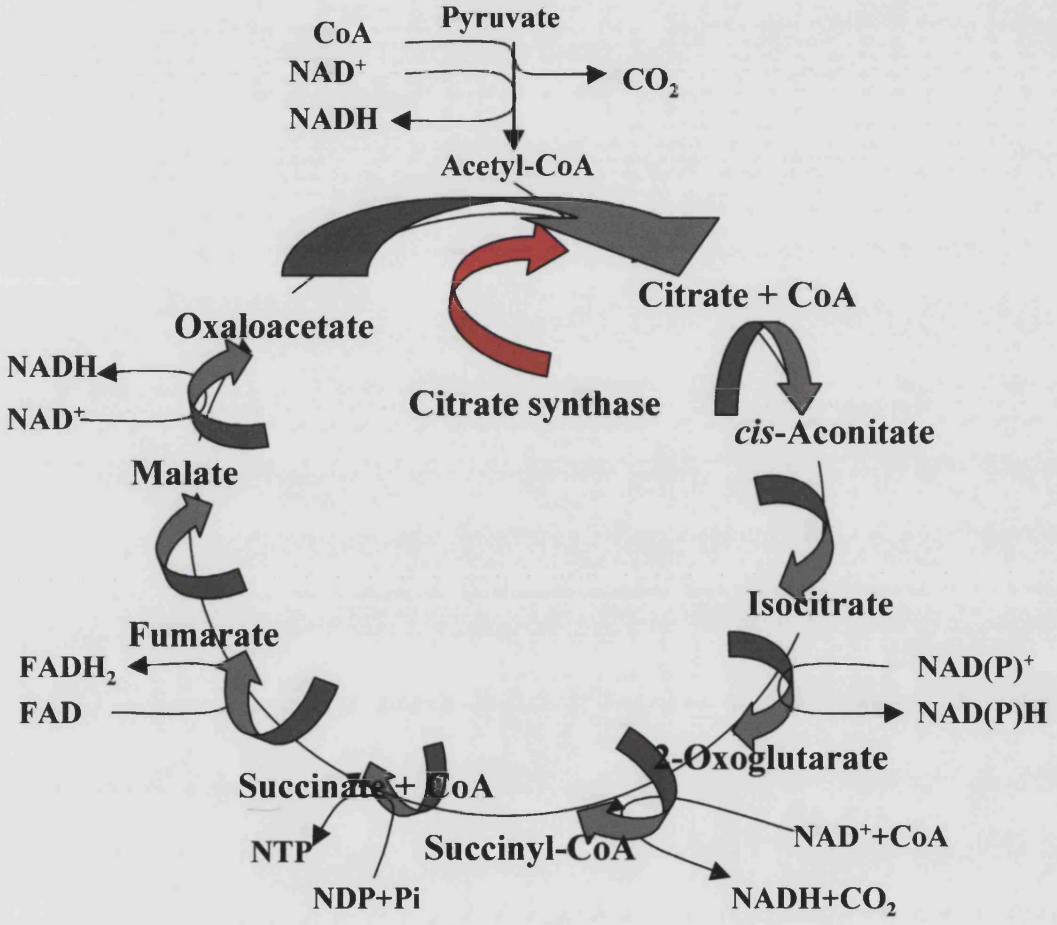


Figure 1.13 Citric Acid Cycle



1.6.2 Structural Forms of Citrate Synthase

The extensive database of structural and mechanistic information available for citrate synthase make it an ideal candidate for comparative enzymology (Russell *et al.*, 1998, Schwartz unpublished, Remington *et al.*, 1982, Russell *et al.*, 1994 and Russell *et al.*, 1997).

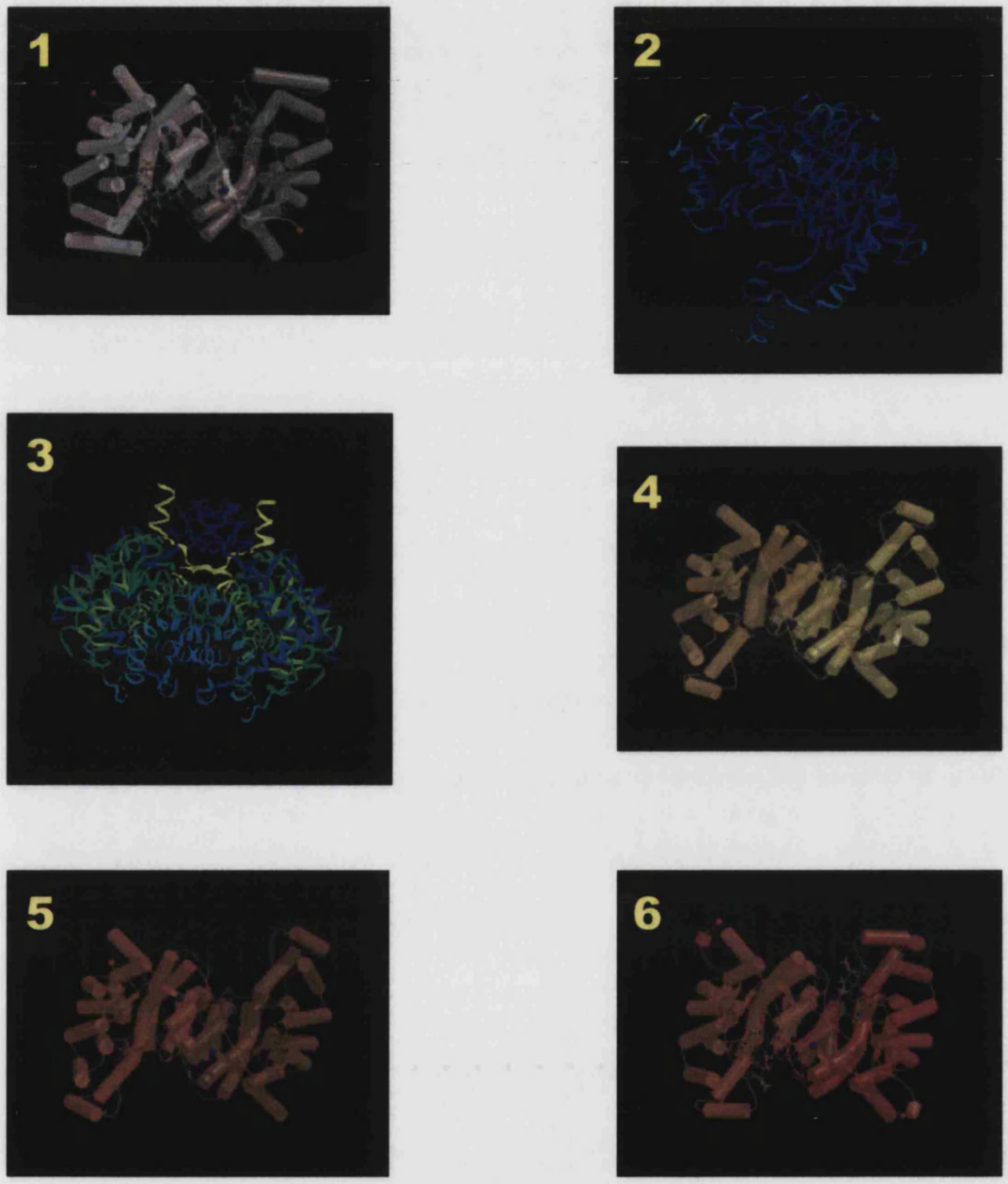
The enzyme has been studied from organisms throughout the three domains, and it has been found in the majority of archaeal phenotypes. Two forms of the enzyme have been discovered, a hexameric and a dimeric form. Both types consist of identical subunits, within the Mr range of 45,000 – 50,000, and the hexamer has been shown to act functionally as a trimer of the dimer (Else *et al.*, 1988).

The hexameric form of the enzyme is found in Gram-negative bacteria, while the dimeric form can be found in Gram-positive bacteria, Eukarya and Archaea (Danson *et al.*, 1985). In both oligomeric forms, each subunit has two domains, a small domain consisting of approximately 5 α-helices and a large domain comprising 11 or more α-helices.

To date there are in excess of 60 primary citrate synthase sequences in the data base and 6 resolved crystal structures of the enzyme from a range of organisms that inhabit a wide span of physical and physiological environments. In the CER at the *University of Bath*, crystal structures have been determined for citrate synthases from extremophilic organisms that span the temperature range at which life is known to exist, that is, the atomic structures of citrate synthase from psychrophilic isolate DS23R (10°C), *Thermoplasma acidophilum* (55°C), *Sulfolobus solfataricus* (85°C), and *Pyrococcus furiosus* (110°C). The structures of citrate synthase from pig (37°C) and chicken (37°C) have also been resolved, these were determined at other institutes. Features that may confer hyperthermostability include greater compactness (shorter loops, lack of cavities, improved packing), an increase in complexes of ion-pairs at the dimer interface and a reduction in thermolabile residues.

It would be of great interest to add to this compilation of crystal structures one from a halophile, and it is this long-term objective that forms the basis of the work reported in this thesis.

Figure 1.14 The six Resolved Crystal Structures of Citrate Synthase



1 Psychrophilic isolate DS2-3R (10°C) (Russell *et al.*,1998) **2** Chicken (37°C) (Schwartz *et al.*, 1999 in press) **3** Pig (37°C) (Remington *et al.*, 1982) **4** *Thermoplasma acidophilum* (55°C) (Russell *et al.*, 1994) **5** *Sulfolobus solfataricus* (85°C) (Russell *et al.*, unpublished) **6** *Pyrococcus furiosus* (110°C) (Russel *et al.*, 1997)

All structures are displayed as homodimers except for 2 Chicken which is displayed as a monomer.

1.7 Project Aims

Work had been conducted over a nine year period to isolate and clone the citrate synthase gene from *Haloferax* (James, 1994). A technique using affinity chromatography which would purify citrate synthase to homogeneity in a one-step process was developed (James *et al.*, 1994), and an N-terminal sequenece, shown to be that of citrate synthase was determined. In 1995 a breakthrough was made when Seedhouse managed to amplify and clone a 800bp gene fragment which was confirmed by sequence comparison to be a fragment of the citrate synthase gene (Seedhouse, BC4 Project Report).

The aim of this project has been to clone, sequence and express the complete citrate synthase gene from the halophilic archaeon *Haloferax volcanii*. High level expression of the halophilic protein was aimed at, using either the in house homologous pBAP5009 halophilic expression system implemented by Jolley *et al.*, (1997) or using a non-halophilic host such as *E.coli*. Using the most suited expression system, sufficient protein was to be purified to enable crystallisation trials to be undertaken, with the long-term aim of resolving a three-dimensional crystal structure for the enzyme.

Prior to this a computer-generated molecular model of the structure was to be generated, based on a known citrate synthase crystal structure with a high sequence identity to the halophilic sequence. This model could then be used as the basis for site-directed mutagenesis studies to explore particular features that may play a role in conferring halphilicity. The expression system used previously to express the halophilic protein would be used to express the mutant enzymes. The information gained from this would be compared with other knowledge already gleaned from the few halophilic structures known to date to gain further insights in the molecular basis of protein halophilicity.

Chapter 2

Materials & Methods

Materials

2.1 General Laboratory Reagents

2.1.1 Cell Culture

Yeast extract, tryptone and bacto-agar were supplied by Difco Laboratories, MI, USA. Bacteriological peptone was obtained from Oxoid, UK. Ampicillin (sodium salt), carbenicillin, tetracycline, novobiocin and chloroamphenicol were purchased from the Sigma-Aldrich Chemical Company Ltd., Poole, UK.

2.1.2 Molecular Biology

All enzymes were supplied by either New England Biolabs (NEB), Hitchin, UK or Boehringer Mannheim, Lewes, UK. 1 kb ladder DNA markers were obtained from Gibco BRL Life Technologies Ltd., Paisley, UK. λ digested with HindIII & EcoRI, used as a molecular weight marker, was purchased from Boehringer Mannheim. Phenol:chloroform:isoamyl alcohol (25:24:1) was obtained from the Sigma-Aldrich Chemical Co., Ltd. Absolute ethanol was purchased from Hayman Ltd., Witham, UK, and water-saturated phenol from Rathburn, Walkerburn, UK. Magic™ Miniprep kits were supplied by Promega Corp., Southampton, UK, SNAP™ Miniprep kits were purchased from Invitrogen, UK and QIAprep™ spin columns, QIAquick™ PCR purification kits and QIAfilter™ maxi kits were acquired from QIAGEN Ltd., Dorking, UK. Geneclean® III kits were obtained from Bio 101, distributed by Anachem, Luton, UK. IPTG and X-Gal were purchased from Calibochem, UK. Ultrapure dNTP sets were obtained from Pharmacia Biotech, St Albans, UK. [α-^{32}P] dCTP Redivue , Hybond-N^{+} nitrocellulose membranes and X-Ray film were from Amersham International Plc., Little Chalfont, UK. X-Ray film was also purchased from Fuji Photo Film Co., Japan, and Blue sensitive X-Ray film was obtained from Genetic Research Instrumentation, Dunmow, UK.

Whatman 3mm Chromatography paper came from Whatman International Ltd., Maidstone, Kent.

2.1.3 Chromatography

Matregel Red A was purchased from Amicon Internationl, UK. Sephadeg-50 was purchased from Pharmacia, UK.

2.1.4 Enzyme Assays

Oxaloacetate was obtained from the Sigma-Aldrich Chemical Co. Ltd., coenzyme A was acquired from Calibochem, UK, and DTNB was brought from Fisons, Loughbourgh, UK.

2.1.5 Electrophoresis

Sea-kelm agarose was purchased from Bio-Rad Laboratories Ltd., Hemel Hempstead, UK. Acrylamide Protogel™ was obtained from National Diagnostics (USA). TEMED, agarose (standard and low melting point) and ethidium bromide came from the Sigma-Aldrich Chemical Co. Ltd., UK. Protein low molecular weight standards were purchased from Bio-Rad Laboratories, UK.

2.1.6 Biochemical Components

All other biochemical components, unless otherwise stated (SLR, AnalaR and Aristar grades), were purchased from BDH Ltd., UK, Sigma-Aldrich Chemical Co., Ltd., Poole, UK and Fisons, Loughborough, UK.

DNA and amino acid sequence analyses were carried out using the GCG Sequence Analysis Software Package Version 7.0 Molecular visualisation was conducted using Rasmol Version 2.5 (PC) written by R.Sayle (Glaxo-Welcome Group Research).

Table 2.1 Escherichia coli Strains

STRAIN	GENOTYPE	SUPPLIER
BL21	*F- ompT hsdS$_B$(r_B-m_B-) gal dcm (DE3)*	Novagen Inc., Madison, USA.
JM109	*RecA1 endA1 gyrA96 thi hsdR17($r_K^-m_K^+$) supE44 relA1 Δ(lac-proAB) [F' traD36 proAB lacI^{q}ZΔM15]*	Promega Corp., Southampton, UK.
XL1-Blue	*RecA1 endA1 gyrA96 thi-1 hsdR17($r_K^-m_K^+$) supE44 relA1 lac* [F' *proAB lacI^{q}ZΔM15* Tn*10* (Tet^r)]	Stratagene Ltd., Cambridge, U.K.
XL1-Blue MRA (P2)	*Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 gyrA96 thi-1 supE44 relA1 lac (P2 lysogen)*	Stratagene Ltd., Cambridge, U.K.

Table 2.2 Haloferax volcanii Strains

STRAIN	SUPPLIER
H.volcanii WFDII	Courtesy of Dr. M. Dyall-Smith University of Melbourne, Australia
H.volcanii New improved strain for transformation	Courtesy of Dr. M. Dyall-Smith University of Melbourne, Australia

Table 2.3 Escherichia coli Plasmids/Haloferax Shuttle Vectors

PLASMID	GENOTYPE & SIZE	SUPPLIER
pGEM®-T	3 kb Amp^r	Promega Corp., Southampton, UK.
pBC SK+/-	3.4 kb ChAmp^r	Stratagene Ltd., Cambridge, UK.
pBAP5009	11 kb Amp^r/ Nov^r	Courtesy of Dr. Keith Jolley, University of Bath, UK.
pET3A	4.64 kb Amp^r/Carb^r	Novagene Inc., Madison, USA.

Glycerol stocks [15% (v/v)] were made of all bacterial strains and bacterial strains containing recombinant plasmids; these were stored indefinitely at -70°C. Pure plasmid preparation of recombinant plasmid DNA were stored in aliquots at -20°C.

2.1.7 Enzymes

Restriction endonucleases and associated buffers, with the exception of *Tha I,* were purchased from New England Biolabs (NEB), Hitchin, UK. *Tha I* was supplied by Boehringer Mannheim, UK. DNA modifying enzymes and polymerases were obtained either from Boehringer Mannheim, Lewes, UK or from NEB. All enzymes were stored at -20°C or as per manufacturer's recommendations. Associated enzyme buffers and appropriate reaction conditions can be seen in Tables 2.4-2.5.

Table 2.4 Restriction Endonucleases, their Buffers and Reaction Conditions used

RESTRICTION ENDONUCLEASE	CONDITIONS	REACTION BUFFER (X1)	HEAT INACTIVATION
Kpn I	Buffer 1 BSA (100μg/ml) 37°C	10 mM Bis Tris Propane-HCl 10 mM $MgCl_2$ 1 mM Dithiothreitol pH 7.9 @ 25°C	NO
Bbv I *Hind III* *Spe I** *Sph I* *Xho I**	Buffer 2 *BSA (100μg/ml) 37°C	10 mM Tris-HCl 10 mM $MgCl_2$ 50 mM NaCl 1 mM dithiothreitol pH 7.9 @ 25 °C	YES YES YES YES YES
Mlu I *Pst I*	Buffer 3 37 °C	50 mM Tris-HCl 10 mM $MgCl_2$ 100 mM NaCl 1 mM dithiothreitol pH 7.9 @ 25°C	YES YES*80
BamH I	Unique Buffer BSA (100μg/ml) 37°C	150 mM NaCl 10 mM $MgCl_2$ 1 mM dithiothreitol pH 7.9 @ 25°C	YES*80
BssH II	Unique Buffer 50 °C	100 mM NaCl 10 mM Bis Tris Propane-HCl 10 mM $MgCl_2$ 1 mM dithiothreitol pH 7.9 @ 25°C	YES
EcoR I	Unique Buffer 37 °C	50 mM NaCl 100 mM Tris-HCl 10 mM $MgCl_2$ 0.025% Triton X-100 pH 7.5 @ 25°C	YES
Sau3A I	Unique Buffer 37°C	100 mM NaCl 10 mM Bis Tris Propane-HCl 10 mM $MgCl_2$ 1 mM dithiothreitol pH 7 @ 25 °C	YES
Tha I(Isoschizomer Mvn I)	Unique Buffer 37°C	10 mM Tris-HCl 10 mM MgCl2 50 mM NaCl 1 mM Dithioerythritol (DTE)	YES

Incubation at 65°C for 20 min can be used to inactivate an endonuclease reaction that has an optimal incubation temperature of 37°C. The enzymes that require a higher inactivation temperature are indicated by *80 and require 20 min at 80°C for complete inactivation. Double digests were conducted using the buffer as recommended by the manufacturer.

Table 2.5 Polymerases and DNA Modifying Enzymes, their Associated Buffers and Reaction Conditions used

ENZYME	BUFFER (1X)	SUPPLIER
$Vent_R$® DNA Polymerase	10 mM KCl 10 mM $(NH_4)_2SO_4$ 2 mM $MgSO_4$ 0.1% Triton X-100 20 mM Tris-HCl (pH 8.8 @ 25°C)	New England Biolabs (NEB)
Expand™ Long Template PCR System	50 mM Tris-HCl (pH 9.2) 16mM $(NH_4)_2SO_4$ 2 mM $MgCl_2$ + detergents	Boehringer Mannheim
Expand™ High Fidelity PCR System	50 mM Tris-HCl (pH 9.2) 16mM $(NH_4)_2SO_4$ 2.25 mM $MgCl_2$	Boehringer Mannheim
DNase I RNase-free	No Buffer	Boehringer Mannheim
RNase I DNase-free	No Buffer	Boehringer Mannheim
Rapid DNA ligation kit (T4 DNA ligase) @ Room temp	20 mM Tris-HCl 1 mM EDTA 5 mM dithioerythritol 60 mM KCl 50% glycerol (v/v)(pH 7.5 @8°C)	Boehringer Mannheim
Klenow Enzyme (labelling grade)	50 mM KPO_4 1 mM DTE	Boehringer Mannheim
T4 DNA ligase @16 °C	50 mM Tris-HCl (pH 7.5) 10 mM $MgCl_2$ 10 mM dithiothreitol 1 mM ATP 25µg/ml BSA	NEB
β-Agarase I	10 mM Bis-Tris-HCl (pH 6.5) 1 mM Na_2EDTA	NEB
Alkaline Phosphatase (calf Intestinal) (CIP)	100 mM NaCL 50 mM Tris-HCl 10 mM $MgCl_2$ 1 mM dithiothreitol(pH 7 @ 25°C)	NEB
Alkaline phosphatase (Shrimp) (SAP)	25 mM Tris-HCl 1 mM $MgCl_2$ 0.1 mM $ZnCl_2$	Boehringer Mannheim

2.1.8 Media and Solutions

The composition of media and solutions used in the Methods section are given below. Solutions were sterilised either by autoclaving at 120°C, 1.41KPa for 20 min, or by filtering using Millipore 0.22μM syringe filters. All solutions were stored at room temperature in colourless glass or plastic bottles, unless otherwise stated.

Table 2.6 Media

MEDIUM	COMPONENTS
***E.coli* Media**	
Luria-Bertani (LB) Broth	1% (w/v) bacto-tryptone 1% (w/v) NaCl 0.5% (w/v) yeast extract
LB-Agar	LB Broth 1.5% (w/v) agar
LB Top Agar	LB Broth 0.7% (w/v) agarose
Liquid culture medium for recombinant phage (λEMBL3)	LB medium 0.2% maltose 10 mM $MgSO_4$
Halobacterial Media	
Halobacterial Media and Solutions taken from "Protocols for Halobacterial Genetics©" version 2.3 (Dyall-Smith *et al.*, 1995).	
Salt Water (SW) Stock Solution 30% (w/v) based on the fromulas described by Rodriguez-Valera *et al.* (1980) and Torreblance *et al.* (1986).	NaCl 240g (g per 1 Litre) $MgCl_2.6H_2O$ 30g $MgSO_4.7H_2O$ 35g KCl 7g
Modified Growth Medium (MGM) 18% SW	Salt Water (30% stock) 600 ml (per 1 Litre) Pure Water 367 ml Peptone 5g Yeast Extract 1g
18% MGM Agar	1 Litre 18% MGM 15 g Difco Bacto-agar
Minimal Medium 18% SW Based on the defined medium published by Rodriguez-Valera *et al.* (1980)	Salt water (30% stock) 600 ml (per 1 Litre) Pure Water 322 ml Tris-HCl (pH 7.5) 1M 20 ml NH_4Cl (1M) 5 ml K_2HPO_4 buffer (0.5M,pH 7.5) 1 ml Trace elements solution 1 ml Carbon source 0.5%

Table 2.7 Antibiotics used in Media

ANTIBIOTIC	STOCK SOLUTION	FINAL CONCENTRATION
Ampicillin	100 mg/ml in water filter sterilised stored @ -20°C	100μg/ml
Carbenicillin	50 mg/ml in water filter sterilised stored @ -20°C	50 μg/ml
Chloroamphenicol	34 mg/ml in ethanol stored @ -20°C	50 μg/ml
Novobiocin	1mg/ml in water filter sterilised stored @ -20°C	0.3 μg/ml
Tetracycline	5 mg/ml in ethanol stored @ -20°C	50 μg/ml

Table 2.8 General Laboratory Solutions

BUFFER	STOCK COMPOSITION
50X TAE	Tris-acetate 0.2 M (pH 8) EDTA 50 mM
10X TBE	Tris-HCl 0.89M (pH 8) H_3BO_3 0.89M EDTA 20mM
TE	Tris-HCl 10mM (pH 7.6) EDTA 1mM
Ethidium Bromide	10mg/ml stock, stored in the dark at 4°C
0.5M Potassium Phosphate	K_2HPO_4 (0.5M) 100 ml KH_2PO_4 (0.5M) 30 ml (+2 ml volumes of KH_2PO_4 until pH reaches 7.5).
Trace Elements sol^n.	$MnCl_2.4H_2O$ 36 mg (per 100 ml) $ZnSO_4.7H_2O$ 44 mg $FeSO_4.7H_2O$ 330 mg $CuSO_4.5H_2O$ 5 mg
IPTG	0.1M stock in H_2O, filter sterilized stored @ -20°C.
EDTA	0.5 M stock in H_2O
6X loading Buffer (2 dye front)	Sucrose 40% Bromophenol blue 0.25% (w/v) Xylene cyanol 0.25% (w/v)
Loading Buffer	Sucrose 45% (w/v) Tris-HCl 0.1M pH 8 EDTA, 50 mM Bromophenol blue 0.25% (w/v)
50X Denhardts Reagent	1% (w/v) Ficoll 1% (w/v) polyvinylpyrrolidine 1% (w/v) BSA
λ/Hind/Eco mw marker (50 ng/μl)	Pre-prepared digested DNA 200 μl Loading buffer 300 μl 50 X TAE 20 μl H_2O 480 μl
Tris-HCl (pH8)	1 M stock, pH with conc NaOH

Table 2.9 Solutions used in Plasmid DNA Preparation

BUFFER	COMPOSITION
Resuspension Buffer	Tris_HCl 25 mM (pH8) EDTA 10 mM Glucose 50 mM
Halophilic Resuspension Bufer	NaCl 1M Tris_HCl 25 mM (pH8) EDTA 10 mM Glucose 50 mM
Lysis Buffer	NaOH 0.2N SDS 1%
Precipitation Buffer	KAc (pH 4.8) 3M
LiCl	6M
chloroform	24:1 (v/v) chloroform:isoamyl alcohol
Phenol	Obtained distilled and stored at 4°C, equilibrated with 2 volumes 50mM Tris-HCl (pH8.0) (stored in the dark at -20°C)
Phenol/chloroform	50% (v/v) Phenol equilibrated with 10mM Tris-HCl, pH7.6 49% (v/v) chloroform 1% (v/v) isoamyl alcohol (stored in the dark at 4°C)

Table 2.10 Solutions used in Southern Blotting

REAGENT	COMPOSITION
20X SSC Stock Sol^{n}. (pH 7)	3M NaCl 175g (g per Litre) 0.3M Na_3 citrate 88g
SM Buffer	NaCl 5.8g (per Litre) $MgSO_4.7H_2O$ 2g Tris-HCl (1M) pH7.5 50 ml Gelatin 2% (w/v) 5 ml
Denaturation Sol^{n}.	1.5M NaCl 44g (g per ½ Litre) 0.5N NaOH 10g
Neutralisation Sol^{n}.	1M Tris-HCl 60g (g per ½ Litre) 1.5M NaCl 44g
Prehybridisation Sol^{n}.	0.5% Blocking Agent 0.1% SDS 5% Dextran Sulfate
Post Hybridisation Wash 1	1X SSC 0.5% SDS
Post Hybridisation Wash 2	0.5X SSC 0.5% SDS
Nylon membrane stripping buffer	1 mM Tris-HCl (pH8) 1 mM EDTA (pH8) 0.1 X Denhardt's reagent
Salmon Sperm DNA	Stock Sol^{n}. 10 mg/ml Working Conc^{n}. 100 µg/ml

Blocking agent was supplied by Amersham International Plc., Little Chalfont, UK, Dextran Sulfate was purchased from United States Biochemical (USB), Cleveland, OH, USA, and Salmon sperm DNA was supplied in molecular biology grade by the Sigma-Aldrich Chemical Co., Ltd., Poole, UK.

Table 2.11 Oligonucleotide Primers and their Sequences used, or Referred, to in the Project

PRIMER NAME	PRIMER SEQUENCE	CHAPTER REF.
λEMBL3L	5′ CTg Atg CCA Tgg TgT CCg ACT TAT 3′	Chapter 3
λEMBL3R	5′ GTT Cag TAA TgA ACC TCT ggA gA 3′	Chapter 3
COSAS	5′ TCC gTg CCA Cag gAg ATA gAg TAC T 3′	Chapter 3
COSS	5′ TAC TCC gCg TCg gAC gTA CTA CCA g 3′	Chapter 3
DM2	5′ TTC gAg AgA ACT Acg CCg T 3′	Chapter 3
DM3	5′ AAT Cgg CAT CgT CCT gTC gTT CAC 3′	Chapter 3
DM4	5′ gTg AAC gAC Agg Acg ATg CCg ATT 3′	Chapter 3
DM7	5′ ATg AAT CCg gAg gAC ATC gAg 3′	Chapter 3
DM8	5′ gAA CTC gCC gAA CAg gAC gAg T 3′	Chapter 3
DM9	5′ ATC Tgg TAg TAC gTC gAC gC 3′	Chapter 3
M13	5′ CgC Cag ggT TTT CCC AgT CAC gAC 3′	Chapter3, 6
M13 reverse	5′ AgCggA TAA CAA TTT CAC ACA ggA 3′	Chapter 3,6
CSBAM	5′ gAT ATg gTA ggA TCC ggT CTg 3′	Chapter 6
CSKPN	5′ TgC ggT ACC TCg CCT ATC gCT 3′	Chapter 6
CSBAMUS	5′ TAT ggC ggT Cgg TAA CAg ATg A 3′	Chapter 6
HCforward	5′ gTC TgC ATA TgT Cag gCC AAC T 3′	Chapter 6
HCreverse	5′ gTg Cgg ATC CTC gCC TAT CgC T 3′	Chapter 6
E86A	5′ TgT TCg gCg AgC gCg CgT gCC Acg T 3′	Chapter 7
E86D	5′ TgT TCg gCg AgA TCg CgT gCC Acg T 3′	Chapter 7
E86Q	5′ TgT TCg gCg AgC Tgg CgT gCC Acg T 3′	Chapter 7
E86S	5′ TgT TCg gCg Agg CTg CgT gCC Acg T 3′	Chapter 7
SDMMLU	5′ Agg AAC TCg Acg CgT TCT CCg A 3′	Chapter 7
SDMXHO	5′ TTg CCT TCT CgA ggT TgA CCT C 3′	Chapter 7

All primers, with the exception of the M13 universal primers and COSS & COSAS, were purchased from Perkin Elmer-Applied Biosystems, Warrington, UK. The M13 universal oligos were obtained from Pharmacia Biotech, St Albans, UK, and primers COSAS & COSS were brought from R+D Systems Europe Ltd., Abingdon, UK.

2.1.9 Buffers used in Protein Analysis Techniques

Table 2.12 Buffers used in SDS-PAGE Gel Electrophoresis

BUFFER/SOLUTION	COMPOSITION
Running Gel Buffer	Tris-HCl 1.5M SDS 0.4% (w/v) pH to 8.9 with Conc HCl, then add SDS
Stacking Gel Buffer	Tris-HCl 0.48M SDS 0.4% (w/v) pH to 6.8 with Conc HCl, then add SDS
Tank Buffer (10X)	Tris-HCl 0.052M SDS 0.1% (w/v) Glycine 4% (w/v)
Load Buffer (2X)	Tris-HCl 0.0125M SDS 4% (w/v) Sucrose 20% (w/v) pH to 6.8 with HCl before adding Bromophenol Blue 0.08% (w/v) Bromophenol Blue 10 % (w/v) β-Mercaptoethanol
Coomassie Staining Soln.	Coomassie Blue R 1.25g Methanol 227 ml H_2O 227 ml Acetic Acid 46 ml
Destain soln.	Acetic Acid 75 ml Methanol 50 ml H_2O 874 ml

Table 2.13 Composition of Various Percentage SDS-PAGE Electrophoresis Gels

RUNNING GEL	2.5%	5.0%	7.5%	8.0%	10.0%	12.5%	STACKING GEL
30%(w/v) Acrylamide (ml)	1	2	3	3.25	4	5	0.9
Gel Buffer (ml)	3	3	3	3	3	3	2.4
H_2O (ml)	8	7	6	5.75	5	4	3.6
0.1 M AMPS (μl)	50	50	50	50	50	50	50
TEMED (μl)	12.5	12.5	12.5	12.5	12.5	12.5	12.5

Table 2.14 Buffers used in Matrex Gel Red A Affinity Chromatography

BUFFER	COMPOSITION
Halophilic Cell cracking Buffer	Tris-HCl (pH 8) 20 mM EDTA 2 mM KCl 2M
Dilution Buffer	Tris-HCl (pH 8) 20 mM EDTA 2 mM
Wash/Equilibrium Buffer	Tris-HCl (pH 8) 20 mM EDTA 2 mM KCl 0.2M
Elution Buffer	0.2 mM Co Enzyme A 1 mM OAA

Table 2.15 Buffers used in Halophilic Citrate Synthase Assays

SOLUTION	STOCK COMPOSITION
$KHCO_3$	**1 M**
Acetyl–CoA	**7 mM** (prepared as described) CoA (Lithium salt) 10 mg KHCO3 (1M) 0.2 ml Acetic anhydride 5µl
DTNB	**10 mM** DTNB 39.6 mg/10 ml Tris-HCl (pH 8) 0.1 M
OAA (freshly prepared)	**10 mM** 13.2 mg/10 ml assay buffer
Halophilic Assay Buffer	Tris-HCl (pH 8) 10 mM EDTA 1 mM KCl 2M

Table 2.16 Solutions used in Transformation of Haloferax Cells

SOLUTION	COMPOSITION
Unbuffered Spheroplasting Sol^n.	NaCl 1M KCl 27 mM Sucrose 15% (w/v)
Buffered spheroplasting Sol^n.	NaCl 1M KCl 27 mM Tris-HCl (pH8.2) 50 mM Sucrose 15% (w/v) [Glycerol 15% (v/v)]
60% PEG_{600} Sol^n.	PEG_{600} 60% Unbuffered spheroplast Sol^n. 40%
Transformation top agar	18% MGM Sucrose 15% (w/v) Agar (6g/L)

Methods

2.2 Strains and Culture Conditions

For exact composition of media please refer to Table 2.6.

2.2.1 Haloferax Strains

Haloferax volcanii WFD II, a derivative of *H. volcanii* DS2 cured of the endogenous plasmid pHV2 (Charlenbois *et al.*, 1987), and strains derived from it were grown in 18% (w/v) salt water modified growth medium (MGM) at 37°C with shaking at 200 rpm. Minimal medium was also used for certain growth experiments; this consisted of the same salts as modified growth medium supplemented with a chosen carbon source and additional trace elements. When required, the medium was solidified by the addition of 1.5% (w/v) agar or 0.6% agar or 0.7% agarose for top agar. Frozen stocks were maintained at –70°C.

Antibiotic resistance plasmids were maintained or selected for in solid and liquid media by addition of ampicillin, tetracycline or novobiocin (Table 2.7)

2.2.2 E.coli Strains

All *E.coli* strains were grown in LB (Luria-Bertani) broth (Sambrook *et al.*, 1989). The medium was supplemented with various antibiotics depending on the genotype of the *E.coli* strain and the presence of a recombinant plasmid. All strains were grown with constant shaking at 200 rpm at 37°C. Media for *E.coli* strains JM109 and XL1-Blue were supplemented with ampicillin (100 μg/ml), for BL21 with either ampicillin or the preferred carbenicillin (50 μg/ml) for longer stability, and for XL1-Blue MRA (P2) with maltose and $MgSO_4$. When required, the medium was solidified by addition of 1.5% (w/v) agar or 0.6% agar or 0.7% agarose for top agar. Frozen stocks were maintained at -70°C. (For genotypes of *E.coli* strains refer to Table 2.1).

2.3 Molecular Biological Methods

2.3.1 Quantitation of DNA

DNA concentrations were determined either spectrophotometrically by measurement of absorbance at 260nm and 280nm, or by comparison of electrophoresed band intensity with a known concentration molecular weight marker on an agarose gel.

An absorbance of 1 at 260nm corresponds to approximately 37μg (single stranded) or 50μg (double stranded) DNA. An A260 / A280 ratio of > 1.8 was taken to indicate purity (Sambrook *et al.*, 1989)

10 μl of λ molecular weight marker, digested with *HindIII* and *EcoRI* of stock concentration 50 μg/ml, produces a high molecular weight top band of approximately 250ng when run on an agarose gel.

2.3.2 Precipitation of DNA

Precipitation was used to concentrate a DNA sample or perform an exchange of solvent. Precipitation of the DNA sample was performed using three volumes of absolute ethanol (stored at -20°C) after which one tenth volume of 3M sodium acetate, pH 5.2 (0.3M final concentration) was added to the sample. This mixture was incubated at -20°C for at least 30 min and the DNA was recovered by centrifugation at top speed in a microcentrifuge for 10 min. The DNA pellet was washed by adding a small volume (~400 μl in a 1.5 ml Eppendorf tube) of 70% (v/v) ethanol, incubating on ice for 2 min and spinning for 2 min at top speed. The ethanol was removed, the pellet dried under vacuum and then resuspended in the desired volume of buffer.

A 0.6 volume of isopropanol (stored at room temperature) was also used to precipitate DNA; this was the method of choice with many of the manufactured plasmid DNA kits as it reduces the prospect of salt precipitation. The method followed is the same as stated above, substituting 0.6 volume of isopropanol for 3 volumes of absolute ethanol and one tenth volume of sodium acetate.

2.3.3 DNA Preparation Methods

2.3.3.1 Preparation of H.volcanii Genomic DNA

2.3.3.2 Rapid Small-Scale Genomic DNA Preparation

This method yields DNA suitable for Southern blotting, although it does not produce high molecular weight DNA suitable for partial enzyme digests. DNA was produced in accordance with the method of Dyall-Smith *et al.* (1995).

200 µl of stationary phase culture was centrifuged for 2 min at 10,000 g in a 1.5 ml microfuge tube. The supernatant was removed with a micropipette. 200 µl of distilled water was squirted onto the cell pellet and rapidly vortexed to disrupt and lyse the cells completely. 200 µl of buffer–saturated phenol (pH8) was added and mixed to extract the protein and carbohydrate from the lysate. This was incubated at 65°C for 10 min, and centrifuged at 10,000 g for 5 min. The aqueous phase (top) was carefully extracted, avoiding any interfacial material, and placed in a clean tube. 400 µl cold absolute ethanol was added, mixed and left to stand on ice for 10 min to allow the DNA to precipitate. The precipitate was centrifuged at 10,000 g for 10 min, the supernatant removed and the remaining pellet washed in 1 ml of 70 % ethanol, which was subsequently removed and the pellet dissolved in 200 µl water. The DNA was stored at –20°C.

2.3.3.3 Large Scale Preparation of Genomic DNA

Genomic DNA was prepared by scaling up and modifying the rapid small scale preparation described above.

1 Litre of stationary phase cells grown in 18% MGM were harvested by centrifugation. The supernatant was discarded and the cell pellet resuspended in 200ml milli Q water and left at room temperature until complete cell lysis had taken place (approximately 20 min). Protein was extracted from the lysate by mixing with 60ml of buffer-saturated phenol (pH 8)

and centrifuging for 10 min (3000 g). A further protein extraction was performed by mixing the aqueous phase with 60 ml of phenol : chloroform : isoamyl alcohol (25:24:1) and centrifuging for 10 min (3000g). If a large amount of protein was present at the phenol aqueous interface this step was repeated. The aqueous layer was removed and the DNA precipitated with either 3 volumes of ethanol, 0.1 volume of sodium acetate or 0.6 volume of isopropanol and 0.3 volume of ammonium acetate. DNA was washed with 70% ethanol and resuspended in a suitable volume of milli-Q Water.

2.3.4 Plasmid Purification

The solutions used in plasmid purification can be found in Table 2.9.

2.3.4.1 Small-Scale Isolation of E.coli Plasmid DNA

Plasmids isolated from *E.coli* were prepared using either Wizard™ minipreps, Qiagen plasmid minipreps or SNAP™ minipreps, or using an adaptation of the alkaline lysis method (Sambrook *et al.*, 1989) depending on what the DNA was going to be used for. Plasmid DNA extracted using the manufactured kits was prepared in accordance with the individual manufacturers instructions. Plasmid DNA prepared using the adaptation of the alkaline lysis method was prepared as follows.

1.5ml of cells from a 10ml overnight culture were centrifuged (12,000g; 1 min). The supernatant was removed by aspiration and the cell pellet resuspended in 100μl of ice-cold miniprep resuspension buffer. After 5 min incubation at room temperature, 200 μl of freshly prepared lysis buffer was added and left for 5 min on ice. A further 5 min incubation was performed after addition of 150μl ice cold precipitation buffer. The precipitate was centrifuged (12,000g; 5 min), and the supernatant removed and extracted with an equal volume of TE-saturated phenol : chloroform : isoamyl alcohol (25:24:1). After a 5 min centrifugation this step was repeated using the aqueous phase and an equal volume of

chloroform : isoamyl alcohol (24:1). DNA was removed using ethanol precipitation, the DNA washed with 70% ethanol and resuspended in a suitable volume of milli Q water.

2.3.4.2 Large-Scale Isolation of E.coli Plasmid DNA

Large scale isolation of plasmid DNA was performed from the appropriate volume of culture. Plasmids from a 100ml *E.coli* culture were prepared using the Qiagen™ Midi system, whereas plasmid DNA from 500 ml cultures was prepared using the QIAfilter 500 Maxiprep system; both methods were carried out in accordance with the manufacturers' instructions.

Plasmid DNA from a 100 ml culture was also prepared as follows. 100ml of cells from an overnight culture were harvested by centrifugation at 6,000 g in a Sorvall centrifuge for 10 min. The cells were resuspended in 4ml resuspension buffer and lysed with the addition of 8ml of freshly-prepared lysis buffer. After standing on ice for 10 min, 6ml of precipitation buffer was added and the mixture left on ice for 15 - 30 min. The resulting precipitate was removed by bench top centrifugation, (3500g ;15 min). DNA/RNA was precipitated from the supernatant by addition of 18ml isopropanol followed by incubation on ice for 30 min. The precipitate was removed by centrifugation and resuspended in 1ml TE. An equal volume of 6M LiCl was added and the mixture was left at room temperature for 15 min. The precipitate was removed and the DNA extracted by ethanol precipitation followed by phenol extraction. DNA was resuspended in 100μl milli Q water.

2.3.4.3 Preparation of H.volcanii Plasmid DNA

Plasmids isolated from *Hf.volcanii* were prepared using the SNAP™ plasmid DNA miniprep kit in accordance with the manufacturers' instructions, or using the method outlined for the small scale preparation of *E.coli* plasmid DNA with the exception that the cells were initially resuspended in a slightly different resuspension buffer. (Please refer to Table 2.1.7 *"Solutions used in plasmid DNA preparation"*).

2.3.5 Agarose Gel Electrophoresis of DNA

DNA fragments were routinely analysed and separated by horizontal agarose gel electrophoresis, using 0.6-1% gels. The gels were made by dissolving agarose in the appropriate volume of TAE buffer by heating in a microwave oven. Ethidium bromide was added to a final concentration of 0.5 μg/ml and the hot agarose poured into a perspegel mould, with a comb in place to form wells, and allowed to set at room temperature for 15 min. Once set, the comb was removed and the gel placed in an electrophoresis tank and covered in TAE buffer. DNA samples were mixed with either loading buffer or the two dye front 6X loading buffer (see Table 2.1.6 *"General solutions"*) and pipetted into the wells. Samples were run in conjunction with a pre-sized digested λ marker at a constant 50-90 V. In certain cases, when DNA bands were to be extracted from the gel and purified, ultra pure low melting point agarose was used. The size separated DNA was visualised under UV transillumination.

2.3.6 Recovery and Purification of DNA from Agarose Gels

2.3.6.1 DNA Purification using Geneclean® III

The Geneclean® III method uses a silica-based matrix called Glassmilk® that has a high affinity for DNA. DNA bands from agarose gels were cut out and DNA extracted and purified in accordance with the manufacturer's instructions.

2.3.6.2 Purification of DNA using PEG 8000

DNA to be purified was precipitated with ethanol, washed with 70% ethanol and air dried. The DNA pellet was dissolved in 32μl of milli Q water and reprecipitated with the addition of 8μl of 4M NaCl followed by 40μl of 13% PEG 8000. The reactants were incubated on ice for 20 min and the DNA pelleted by centrifugation (12,000 g ; 5 min) at 4°C. The supernatant was removed and the pellet rinsed with 500μl of 70% ethanol and dried under vacuum for 3 min.

2.3.6.3 Purification of DNA using β Agarase Digestion

DNA of interest was excised from a low melting point agarose gel. The agarose slice was melted at 65°C for 10 min and 10% total volume of 10x agarase buffer was added and cooled to 40°C before β-agarase was added at a concentration of 2 μl enzyme for every 200μl agarose gel. The reactants were incubated at 40°C for 2 h. 10% (v/v) of 3M NaOH (pH 5.5) was added to the tube and the contents incubated for 15 min on ice. Undigested agarose was removed by centrifugation (13000g ; 15 min) . DNA was precipitated from the supernatant by addition of 2 x volumes of isopropanol.

2.3.7 Restriction Digestion of DNA

DNA was digested with the appropriate restriction endonuclease as specified in the manufacturer's instructions. The DNA was incubated with the enzyme in the specified buffer (supplied as a 10X concentrate) at 37°C, unless otherwise stated. In most cases of double digestion, the enzymes could be used simultaneously referring to the manufacturer's instructions for the correct buffer to use. However, if the individual buffer requirements were not suited to the reciprocal endonuclease, the digestion was performed sequentially, digesting first with the enzyme requiring the lowest buffer concentration. The buffer concentration was then adjusted to suit the second enzyme. Alternatively, the DNA was ethanol precipitated after the first incubation and resuspended in the second reaction buffer. Incubation times varied from 30 min for a partial digestion to overnight.

2.3.8 Dephosphorylation of Linearized DNA

The digested DNA to be dephosphorylated was extracted with phenol:chloroform and ethanol precipitated. The DNA was recovered by centrifugation and redissolved in 90 μl of 10 mM Tris-HCl (pH 8.3). To this, 10 μl of the 10X dephosphorylation buffer and the appropriate amount of alkaline phosphatase were added. To dephosphorylate protruding 5′ termini, dephosphorylase was added to a final concentration of 1 unit/ 100 pmoles and incubated for 30 min at 37°C. For blunt or recessed termini, the enzyme was supplied at an initial concentration of 1 unit/2 pmoles and incubated at 37°C for 15 min. A further aliquot of dephosphorylase was then added and a further incubation at 55°C for 45 min took place. At the end of the incubation period, 0.5% (v/v) SDS, 5 mM EDTA (pH8), and protinase K to a final concentration of 100 μg/ml were added and incubated at 56°C for 30 min. The reaction was cooled to room temperature when the DNA was extracted with phenol, phenol:chloroform

and precipitated with ethanol. After washing the DNA with 70% ethanol it was redissolved in an appropriate volume of water.

Dephosphorylation with calf intestinal phosphatase (CIP) requires the enzyme to be completely removed for subsequent ligations to work efficiently. However, dephosphorylation with shrimp alkaline phosphatse (SIP) does not require the complete removal of the enzyme and the DNA can be precipitated directly after incubation with the enzyme.

2.3.9 Ligations

Ligation of cohesive ends was carried out using 1 unit of T4 DNA ligase (supplied at a concentration of 1 unit/μl) in the buffer supplied with the enzyme. A ratio of insert to vector of 1-6 : 1 was used in a total volume of 30 μl. The reaction was incubated overnight at 15°C or at room temperature for 3h. 15 μl of the reaction mixture was used to transform competent *E.coli* cells, with the remainder kept at 4 °C for use if the transformation failed.

Alternatively, fast acting T4 DNA ligase was used, for which the ligation was set up in accordance with the manufacturer's instructions and the reaction took place at room temperature for 5 min. 1 μl of the reaction mix was used used to transform competent *E.coli* cells.

2.3.10 Construction of Plasmids

Plasmid DNA and insertional DNA were digested with restriction endonucleases. Fragments were isolated by agarose-gel electrophoresis, and DNA was recovered from the agarose using one of the above-mentioned DNA recovery methods. DNA from single enzyme digests was treated with alkaline phosphatase as directed above. However, it was assumed that double-digested fragments were unlikely to religate to themselves and so they were not treated with either *CIP* or *SAP*. Specific fragments were ligated together using T4 DNA ligase as outlined in section 2.3.9 Ligations.

2.3.11 *Polymerase Chain Reaction (PCR) Amplification of DNA*

All primers used in this thesis are listed in Table 2.11.

Reactions contained approximately 100 ng target DNA, 0.1 μM of each primer, and 50 μM each of dATP, dCTP, dGTP and cTTP in a 100 μl final volume. Magnesium was also present in the reaction, varying in concentration (0.5 - 2.5mM) depending on the polymerase used and the reaction conditions employed. Magnesium was usually supplied in the form of $MgCl_2$ or $MgSO_4$. The reaction mixture was overlaid with 50 μl mineral oil to prevent evaporation. Reactions were incubated at 96°C for 5 min before the addition of 1-2.5 units of polymerase. The amplification program varied, but consisted of a melting step (usually conducted at 96°C for 75 s), an annealing step, the temperature of which depended on the melting point of the various primers employed, and an elongation step, usually conducted at 72°C (length determined by size of expected product, usually 1 min per kilobase). A final 10 min extension at 72 °C was run after the last cycle was complete. The reactions were then maintained at 4°C. A 10 μl aliquot from each reaction was run on a 1% (w/v) agarose gel. PCR reactions were carried out on a Perkin-Elmer /Cetus DNA thermal cycler.

2.3.11.1 Nested PCR of Haloferax volcanii Genomic DNA

Nested PCR was carried out on genomic DNA using the following thermal cycling conditions for denaturation, annealing and elongation, respectively:
1.25 min @ 96°C, 1.5 min @ 50°C and 4 min @ 72°C for 30 cycles. This was followed by a 10 min final extension at 72°C.

2.3.11.2 PCR using Gene Specific Primers

Amplification was carried out on *Haloferax volcanii* genomic DNA using the thermostable polymerase vent®. PCR reactions were carried out using the following conditions for denaturing, annealing and elongation, respectively:
1.25 min @ 96°C, 1.5 min @ 55°C and 2 min @ 72°C for 30 cycles, followed by a 10 min final extension at 72°C.

2.3.11.3 Touchdown PCR

The method of touchdown PCR is designed to minimise non-specific binding of primers and increase the concentration of the required product. It employs a 30-step cycle starting with a high annealing temperature (60°C); it then goes through 2 full cycles before lowering the annealing temp by increments and repeating the cycles. The annealing temperature is lowered by 1° increments every other cycle until it reaches 50°C. The higher the annealing temperature the greater the specificity and the more chance of amplifying the desired product.

Touchdown PCR was carried out on *Hf.volcanii* genomic DNA using the thermostable polymerase vent® under the following thermocycling conditions for denaturation, annealing and elongation, respectively: 1.25 min @ 96°C, 1.5 min @ 60 → 50°C and 2 min @ 72°C for 2 cycles at each annealing temperature. A further 20 cycles were carried out at a constant annealing temperature of 50°C followed by a 10 min final extension at 72°C.

2.3.12 Southern Blotting

Southern blots were carried out with modifications to the protocol described (Sambrook *et al.*, 1989). Solutions used in this process can be found in Table 2.10.

Single and double restriction digests of genomic DNA were carried out with various enzymes as described previously. Digests were run on 1% (w/v) agarose with 0.5 μg/ml ethidium bromide in 1x TAE buffer overnight at 15V (approximately 1V/cm). The gel was denatured and partially hydrolysed prior to DNA transfer to a nylon membrane (Wahl *et al.*, 1979, Meinkoth and Wahl , 1984).

When using large DNA fragments, the DNA was nicked by brief depurination by soaking the gel in several volumes of 0.2N HCl. DNA was denatured by soaking the gel in several volumes of denaturation buffer for 45 min, and neutralised by soaking for 30 min in several volumes of neutralisation buffer. Neutralisation buffer was replaced and the gel soaked for a further 15 min. Between each step the gel was washed with deionised water. Blotting onto Hybond N^+ Nylon membranes was carried out under neutral conditions by the capillary method (Southern, 1975). The Hybond N^+ membrane was prewetted with deionised water, and immersed in transfer buffer (10x SSC) for 5 min; the blotting stack was set up using a reservoir containing 10xSSC. Blotting was carried out for 18 h. The membrane was removed, washed in 6x SSC and air dried. DNA was baked onto the membrane at 80°C for 1 hour or by crosslinking with UV transillumination.

2.3.13 Colony Hybridisation

Gridded Hybond N^+ nylon membranes were placed for 2 min on the surface of LB/agar containing top agar and λ phage plaques. The LB/agar plates had been placed at 4°C for an hour prior to blotting. Hybond N^+ membranes were orientated on the plates by injecting black waterproof ink through the membrane and into the top agar. The membranes were placed (plaque side up) on a pad of Whatmann paper soaked in denaturation buffer for 7 min. They

were then transferred to a fresh pad soaked in neutralisation buffer for 3 min. Denatured membranes were then washed in 2x SSC, air dried for 20 min and the DNA fixed onto the membrane by baking at 80°C for 1 hour or by crosslinking with UV transillumination.

2.3.14 Labelling of Oligonucleotide Probes

Oligonucleotides were labelled by the "random primed" method (Feinberg and Vogelstein, 1983, 1984).

DNA to be labelled was denatured for 10 min at 100°C and cooled on ice. 25 ng denatured DNA was added to the following in a microfuge tube on ice: 100 μM of dATP, dGTP and dTTP, 2 μl hexa-nucleotide mixture in 10x reaction buffer, 5 μl [α -^{32}P] Redivue dCTP (3000 Ci mmol^{-1}) made up to a total volume of 19 μl . Finally 1 μl Klenow enzyme (2units/μl) was added and the reaction incubated for 30 min at 37°C.

Incorporation of radiolabel was checked and unincorporated label removed by passing the solution through a Sephadex ® G-50 spun column centrifuged at 3000rpm for 10 min.

2.3.15 *Hybridisation of Labelled Oligonucleotide Probes to DNA Blots*

Hybridisation of oligonucleotide probes was carried out with modifications to the method described by Sambrook *et al*. (1989).

The dried membranes were placed in a glass hybridisation bottle containing prehybridisation buffer at 0.2ml per cm^2 of membrane (approximately 5ml per colony blot membrane). The solution was incubated at 60 - 65°C for 10 min. Salmon sperm DNA was denatured at 100°C for 10 min and added to the tube to a final concentration of 100 μg /ml.

The membranes were incubated for 2 h at the chosen temperature. Labelled α-^{32}P oligonucleotide probe was denatured at 100°C for 10 min and placed on ice; 10 µl was added to each tube and the membranes incubated for a further 12 - 18 h.

The membranes were washed for 30 min with wash buffer 1 and for a further 30 min with wash buffer 2. Washed membranes were kept moist, wrapped in clingfilm and autoradiographed using X-ray film and intensifying screens at -70°C. Films were developed after 5 hour and 18 hour exposures depending on the intensity of the probe. The film was developed in an Amersham Hyperprocessor.

2.3.16 Removal of Probes from Nylon Membranes

Radiolabelled probes were removed from Hybond N^+ membranes by immersing the membrane in 200 ml of membrane stripping buffer for 2 h at 75°C. The membranes were washed with 2x SSC, air dried and kept in aluminium foil to be reprobed.

2.3.17 Making and Screening a λ EMBL3 Library

2.3.17.1 Preparation of Partially Digested DNA

Small scale Sau3AI digests using various concentrations of enzyme (10 units-0.156 units) were peformed on *Hf. volcanii* gDNA (1-2µg) for 30 min at 37°C. The reaction was stopped by adding 30 µl 0.5M EDTA and 10 µl of the digest was viewed on a 1% agarose gel. Concentrations of enzyme were selected that induced complete DNA digestion in that time period and each reaction was scaled up X5. The above procedure was repeated with the scaled-up digests. Complete DNA digests were chosen that gave the bulk of the fragments in the size range (7 -15 kb) and the various chosen digests were pooled, precipitated with ethanol and resuspended in 500 µl Milli-Q H_2O. The DNA fragments were purified and size selected further by rate zonal centrifugation. The pooled DNA fragments were placed on a

NaCl gradient ranging from 5 - 25 % NaCl and centrifuged at 22,000 rpm in a Sorval ultra-centrifuge for 16 h. Aliquots of 500 µl were collected and 10 µl of each run on a 1% agarose gel. Fractions containing DNA fragments of molecular weight 10 -15 kb were combined, diluted with H_2O, ethanol precipitated, washed with 70% (v/v) ethanol and resuspended in a total volume of 200 µl H_2O.

2.3.17.2 Ligation to λEMBL3 BamHI Predigested Arms

1 µl (1 µg) of predigested λ EMBL3/BamHI was ligated to an approximate equi-molar ratio of size fractionated *Hf. volcanii* DNA (~1 µl). Although the pre-digested λ arms were digested with BamHI, compatible inserts could be digested with BamHI isosizomers Sau3AI, MboI or Bgl II. Lambda EMBL3 vector arms can accommodate inserts ranging from 9-23 kb. Ligations took place using 1 unit of T4 DNA ligase in a total reaction volume of 10 µl at 15°C overnight.

2.3.17.3 Packaging Ligated λEMBL3 Clones into Phage Heads

Lambda clones were packaged using Gigapack® III gold packaging extract, which is an in vitro packaging extract that preferentially size selects for extra large inserts, while still maintaining high packaging efficiencies. 2.5 µl of the ligation mixture was mixed with 25 µl of the extract mix and incubated at room temperature for 2 h. Highest packaging efficiencies have been shown to occur between 90 min and 2 h, efficiency dropping dramatically after this time period. After the incubation time, 500 µl SM buffer and 20 µl chloroform were added and the mixture stored at 4°C.

2.3.17.4 Preparation of Host Bacteria

A glycerol stock of *E.coli* XL1-Blue MRA (P2) was streaked onto an LB plate and incubated overnight at 37°C. LB supplemented with $MgSO_4$ and maltose (see Table 2.1.4 "Media") was inoculated with a single colony. This was grown at 37°C, shaking at 200 rpm to an OD_{600} =1. The bacteria were pelleted by centrifuging at 2000 rpm for 10 min. The cells were resuspended in half the original volume with sterile 10 mM $MgSO_4$ and diluted to OD_{600} = 0.5 with 10 mM $MgSO_4$.

2.3.17.5 Diluting the λEMBL3 Hf volcanii Library

Host bacteria were prepared as outlined above. Dilutions of the final packaged reaction mixture were made in SM buffer in the ratios 1 in 10 to 1 in 10^6. 1 µl of the undiluted packaged mix was added to 200 µl host cells diluted in 10 mM $MgSO_4$ to OD_{600} = 0.5. For dilutions of the package mix, 10 µl of the dilution was added to 200 µl host cells. The phage and the bacteria were incubated for 15 min at 37°C with gentle shaking. 4 ml top agar (<50°C) was added to the cells and this was then plated onto LB-agar plates and incubated overnight at 37°C. Only recombinant phage will grow on XL1-Blue MRA (P2), and plaques should be visible after 8-12 h of incubation.

2.3.17.6 Amplification of the λEMBL3 Library

Host bacteria were prepared as outlined above. 600 µl of host cells were required per 150 mm plate. 20 µl of a 1 in 100 dilution of the packaged mixture containing approximately 50, 000 plaque-forming bacteriophage were mixed with 600 µl aliquots of prepared host cells. The mixture was incubated at 37°C for 15 min. 7 ml of melted top-agar (<50°C) was added to the mixture and poured on to a 150 mm LB-agar plate. Plates were incubated for 6-8 h at 37°C to produce small plaques of ~1mm. The plates were overlayed with 10 ml SM buffer

and stored overnight at 4°C with gentle rocking. The bacteriophage suspension was recovered from the plates and pooled, the plates were rinsed with an additional 2 ml SM buffer and pooled with the rest of the suspension. Chloroform was added to a final concentration of 5% (v/v) and incubated at room temperature for 15 min. Cell debris was removed by centrifugation and chloroform was added to a final concentration of 0.3% (v/v). The library was stored at 4°C and aliquots were stored in 7% DMSO at -70°C.

For a representative amplification of the library, 20 X 150 mm plates were used per amplification procedure.

2.3.17.7 DNA Screening Protocol

Host cells were prepared as outlined above. 1 in 10 and 1 in 100 dilutions of the original library stock were prepared in SM buffer. 300 µl of host cells were added to 10 µl of the diluted library stock and incubated for 15 min at 37°C. 4 ml of melted top-agar (<50°C) was added to each 300 µl mix of bacteriophage, which was then plated on to 7 mm LB-agar plates. These were incubated overnight at 37°C. Plaques should form as clearings on a bacterial lawn. Plates were then set up at 4°C for 30 min prior to undergoing colony hybridisation (see section 2.3.13 "*Colony hybridisation*"). A typical screening used 10 X 1 in 10 dilution plates and 10 X 1 in 100 dilution plates. Plaques that gave a positive result after screening were removed with a glass pipette and placed in a 1.5 ml Eppendorf tube containing 100 µl SM buffer. This was left at 4°C overnight. 20 µl of this lysate was used to inoculate 300 µl of host cells, which were used to form an LB top-agar plate as described previously. This plate was then used to form a secondary colony hybridisation screen executed in the same manner as the first. If the correct positive plaque was chosen, every plaque which appeared on the plate should give a positive result on X-ray film. If all plaques were positive then the plate was over-layed in SM buffer overnight at 4°C. The buffer was removed and the lysate centrifuged with 5% (v/v) chloroform to remove cell debris; this was then stored indefinitely in a final concentration 0.3% (v/v) chloroform at 4°C. If not every

plaque on the secondary screen plate gave a positive result on X-ray film, then individual positive plaques would be removed with a sterile glass pipette and stored in 100 µl SM buffer/plaque with 0.3% (v/v) chloroform at 4°C.

2.3.17.8 Preparation of a λ Phage Cell Lysate

A fresh culture of *E.coli* XL1-Blue MRA (P2) was started by inoculating a single colony from an agar plate into 5ml of LB supplemented with 0.2% (v/v) maltose and 10 mM $MgSO_4$. This was grown overnight at 37°C. 1.25 ml of the overnight culture was added to a tube containing 50 µl of a top-agar plug eluate, and incubated at 37°C for 30 min. 1.25 ml of the infected culture was transferred into 100 ml of prewarmed (37°C) LB supplemented with 10 mM $MgSO_4$. This was left shaking at 37°C overnight with good aeration. Cell debris should be visible the next morning and the medium should appear clear if cell lysis has occurred. 500 µl chloroform was added and shaken at 37°C for a further 15 min. The cell lysate was centrifuged at 8,000 G for 10 min to remove the cellular debris. The cell lysate can be stored for up to 6 months at 4°C. Aliquots were taken and stored for infection of future cultures. For a 300 ml cell lysate, 4 ml of host cells were combined with 150 µl of phage lysate.

2.3.17.9 Preparation of λ Phage DNA

λ DNA was prepared using QIAGEN® λ Maxi kits. A 300 ml cell lysate was prepared as described above. The DNA was prepared in accordance with the manufacturer's instructions except the extra DNase and RNase were added just after inoculation of the LB culture with bacteriophage.

2.3.18 Sequencing

Sequence analysis was performed using dye-labelled terminator cycle sequencing chemistry on an Applied Biosystems 373A DNA sequencer in accordance with the manufacturer's instructions. The reaction mixture consisted of 250-500 ng double stranded DNA for plasmid sequencing and 750 ng-1 μg DNA for sequencing λ DNA, and 5 pmol of the appropriate primer in a total volume of 6 μl. 4 μl of reaction mixture (Applied Biosystems) was added to this mixture prior to under going thermal cycling.

2.3.19 Site-Directed Mutagenesis

Site-directed mutagenesis was undertaken using PCR. The desired mutations were incorporated into PCR primers, and the region to be mutated was flanked by two unique restriction sites that were also incorporated into PCR primers. PCR reaction 1 was performed using one of the restriction site primers and the primer containing the mutation. This produced a short PCR product incorporating the mutation which was then used as a primer in conjunction with the 2nd restriction site primer in a second PCR reaction. The PCR product was run on a 1% (w/v) agarose gel and purified using one of the above mentioned methods. The region in the WT gene was also restricted with the appropriate endonucleases and the linearized fragment purified in the same way as the PCR product. The mutant fragment was then ligated into the WT gene using rapid T4 DNA ligase. PCR conditions used to make the mutants were 96°C for 75 s, 55°C for 1 min, 72°C for 2 min and a final extension of 72°C for 10 min.

2.3.20 Transformation of E.coli

2.3.20.1 Preparation of Competent Cells

E.coli JM109 or XL1-Blue were streaked out onto LB agar and incubated overnight at 37°C. 50 ml LB was dispensed into a 500 ml flask, 1 ml of this was transferred into a 1.5 ml Eppendorf tube. Several of the overnight colonies were sub-cultured into 1ml LB and then subsequently transferred into the 500 ml flask. The culture was incubated at 37°C, and shaken at 200 rpm until the cells reached an OD_{600} = 0.4. The cells were then pelleted at 5,000 rpm for 5 min and the supernatant removed. The cells were resuspended in 25 ml of ice cold 50 mM $CaCl_2$ and left to stand on ice for 20 min. The cells were repelleted at 5,000 rpm for 5 min and resuspended in 5 ml of 50 mM $CaCl_2$. The cells can be stored on ice for up to 24 h or frozen in the presence of 15% (v/v) glycerol in 200 µl volumes in an acetone-dry ice bath and stored at -70°C.

2.3.20.2 Transformation of JM109 and XL1-Blue using the Heat-Shock Method

200 µl of competent cells were thawed on ice. To these the transforming DNA was added (up to 50 ng in a volume of 10 µl or less), mixed gently and the tubes then stored on ice for 30 min. The tubes were then heat-shocked in a 42°C water bath for 2 min and then immediately put on ice for 2 min. 1 ml of LB was added to each tube and the cultures incubated at 37°C for 1 h to allow the bacteria to recover and express ampicillin resistance. The cells were then spun at 5,000 rpm in a microfuge and the cells resuspended in a total volume of 200 ml supernatant; this was plated on pre-warmed (37°C) LB-agar plates supplemented with ampicillin 100 µg/ml. The plates were incubated overnight at 37°C and transformed colonies selected. (Modified from Sambrook *et al.*, 1989).

2.3.20.3 Transformation of BL21 (DE3) using the Heat-Shock method

1 μl of ligation reaction was added to 20μl of pre-thawed competent cells on ice. The components were gently mixed and left on ice for 30 min. The tubes were then heat shocked for exactly 40 s at 42°C and placed on ice for 2 min. 80 μl LB medium was added to each tube and incubated at 37°C for 1 h. 50 μl of each transformation was spread on LB-agar plates containing carbenicillin 50 μg/ml. The plates incubated at 37°C overnight and transformants were selected.

2.3.21 Transformation of Hf.volcanii

The solutions employed in this method are listed in Table 2.16.

Transformations were peformed using the PEG method described by Charlebois *et al.* (1987). Freshly-inoculated cultures were grown until late log phase (OD_{600} of 0.8-1.0) and were then harvested by centrifugation at 3,300 g for 15 min, at room temperature. The cells were then washed in 1/10 volume buffered spheroplasting solution and resuspended in buffered spheroplasting solution containing 15% (v/v) glycerol. Spheroplasts formed on addition of 45 mM EDTA (pH 8) and they were then added to tubes containing 2-5 μg DNA and an equal volume of 60% (v/v) purified PEG_{600} added. The cells were incubated for 20-30 min at room temperature after which they were allowed to recover in 18% (w/v) MGM overnight. Selection of transformants was by plating onto 18% (w/v) MGM-agar containing novobiocin 0.3 μg/ml and incubating at 37°C for 5-10 days.

2.4 Biochemical Methods

2.4.1 Preparation of Cell Extracts

E.coli cells were lysed in halophilic cell cracking buffer only if they were expressing a halophilic protein. In other cases they were lysed in the same buffer without 2M KCl.

E.coli cells were harvested by centrifugation at 10,800 g for 15 min at 4°C. The cells were then resuspended to a density of 0.2 g/ml in the appropriate cell cracking buffer containing 1 mM PMSF and lysed by sonication on ice in an MSE 150 W ultrasonic Disintigrator Mk2 using a 9.5 mm end diameter titanium probe. Three 30 s bursts were used with a 90 s cooling time between each. Cell debris was removed by centrifugation at 15,600 g for 30 min at 4°C in a Sorvall SS34 rotor.

For small volumes, cells were harvested at full speed in a microcentrifuge for 2 min in 1.5 ml Eppendorf tubes, and resuspended in a reduced volume. The samples were sonicated using a 3 mm end diameter exponential probe. Three 10 s blasts were used with approximately 1 min cooling time between each. Cell debris was removed by further centrifugation at full speed for 10 min.

Hf.volcanii extracts were prepared in a similar manner *to E.coli* cells.

2.4.2 Citrate Synthase Assay

All appropriate assay buffers are listed in Table 2.15.

Citrate synthase activity was assayed spectrophotometrically at 30°C in a Perkin Elmer Lambda Bio or Lambda 11 spectrophotometer by the method of Srere *et al.*,(1963). In this method coenzyme A produced by the action of citrate synthase reacts with DTNB, releasing thionitrobenzoate, which has an absorption maximum at 412 nm and a molar absorption coefficient of 13,600 l mol^{-1} cm^{-1}. Initial rates were calculated using PECSS software supplied with the machine.

Assays were carried out in halophilic assay buffer containing 0.2 mM oxaloacetate, 0.4 mM acetyl CoA and 0.1 mM DTNB in a final volume of 1 ml.

2.4.3 Acetylation of Coenzyme A

Coenzyme A was acetylated by dissolving 10 mg in 1 ml of water, cooling on ice and adding 0.2 ml of 1 M $KHCO_3$. 5 µl of acetic anhydride was added and the solution left on ice for 20 min. Acetylation was tested by reacting with 0.1 mM DTNB and monitoring at 412 nm.

2.4.4 Techniques used in Protein Analysis

The composition of buffers and other components used in this technique are listed in Tables 2.12. and 2.13.

2.4.4.1 Estimation of Protein Concentration

Bradford Assay

Protein concentration was estimated by the method of Bradford (1976) using either a kit from Bio-Rad or Bradford Reagent prepared in the laboratory, and a standard curve prepared from dilutions of bovine serum albumin (0 to 1 mg/ml).

100 µl of appropriately-diluted sample was mixed with 1 ml Bradford Reagent in a cuvette. This was incubated at room temperature for approximately 10 min. After initial colour development, the sample remained stable for up to 1 h. The absorbance at 595 nm was determined and the protein concentration interpolated from known standards.

2.4.4.2 Sodium Dodecyl Sulphate (SDS) Polyacrylamide Gel Electrophoresis

The composition of buffers and other components used in this technique are listed in Tables 2.12 and 2.13.

2.4.4.3 Preparation of Samples

Protein samples were mixed with Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE) Sample Buffer, boiled for 5 min and allowed to cool before analysis by SDS-PAGE. Molecular weight markers, sometimes pre-stained (Bio-Rad), were used according to the manufacturers' instructions.

2.4.4.4 Preparation and Running of Polyacrylamide Gels

Gels were poured and run in an Atto Corp. Mini-Atto system. Resolving gels were made by mixing the components in Table 2.13. Usually 10% acrylamide gels were made, and polymerisation was achieved by the use of ammonium persulphate (APS) and N,N,N',N-tetramethylethylenediamine (TEMED). Gels were overlaid with water-saturated butan-1-ol and allowed to set for 20 min. Stacking gels were made by combining the necessary component listed in Table 2.12. The butan-1-ol layer was discarded and the comb inserted between the gel plates. The stacking gel was poured onto the resolving gel and allowed to set for a further 40 min. 10µl of sample was loaded into each well and gels were run in tank buffer. Gels were run at 10 mA/gel until samples entered the resolving gel and then at 20 mA/gel.

2.4.4.5 Coomassie Stain of Polyacrylamide Gels

Following electrophoresis, the gel was stained for 1 h at room temperature in Coomassie Blue Stain and then destained in several changes of destain until the background colour was removed.

2.4.5 Matrex Gel Red A Affinity Chromatography

A 2 cm column of Amicon Matregel Red A was equilibriated with 10 ml equilibrium buffer. Cell extract, diluted to a final concentration of 0.2M KCl with dilution buffer, was added to the column. This was washed with 10 ml wash buffer containing 0.2 M KCl to remove any unbound protein. The sample was removed from the column with 6 ml elution buffer and the eluant collected in 2 ml fractions. The fractions were tested for citrate synthase activity prior to the column being cleaned with 5M urea and 0.5M NaOH, before being re-equilibriated with equilibrium buffer.

Chapter 3

Cloning and Sequencing of the Citrate Synthase Gene from Haloferax volcanii

3.1. Background to the project

Prior to the start of this project, an investigation of citrate synthase from this organism had already been undertaken. Citrate synthases from both thermophilic and halophilic Archaea had been purified to homogeneity using affinity chromatography on Matrex Gel Red A (James *et al.*, 1994). The determination of the N-terminal amino acid sequence of this enzyme from *Haloferax volcanii* (James, 1994) provided a template on which to base oligonucleotide primer design.

It is known that the *Hf volcanii* genome has a high G+C content (approximately 60%). This is reflected in the genes sequenced from it at this time: dihydrolipoamide dehydrogenase (Vettakkorumakankav *et al.*, 1992), dihydrofolate reductase (Zusman *et al.*, 1989), his C (Conover & Doolittle, 1990) and the trp cluster (Lam *et al.*, 1990). Based upon this information a codon usage table was compiled (James, 1994) which enabled oligonucleotides to be designed.

3.1.1 PCR Amplification of a Citrate Synthase gene fragment

Degenerate PCR primers were designed based on the known N-terminal amino acid sequence and an internal consensus sequence, close to the C-terminus, predicted from multiple sequence alignments of citrate synthase sequences. A nested PCR approach was selected and four oligonucleotides designed. The outer primers CHS1 and CHS4 were used initially to amplify the gene coding for citrate synthase directly from genomic DNA (gDNA), and the inner (nested) primers CHS2 and CHS3 were used to amplify further this initial PCR product. The aim of these reactions was to produce a single product which would correspond to a large percentage of the citrate synthase gene, which in turn could be cloned, labelled and used as a probe to obtain the complete the gene sequence.

3.1.2 Sequencing of the PCR fragment

A single band of the expected size (880 bp) was eventually obtained. The nested PCR product was isolated, purified and cloned into the *SmaI* restriction site in the plasmid pUC18 and transformed into *E.coli* XLI Blue cells. Using universal primers, the clone was then sequenced.

Multiple sequence alignments of this putative *Hf.volcanii* citrate synthase sequence and known citrate synthase sequences from a number of organisms showed that it contained conserved amino acid residues, indicative that it was a citrate synthase fragment. This work was carried out by a final year undergraduate project student and is documented in her project report (Seedhouse, 1995).

3.2 Preparation of Hf volcanii genomic DNA

Genomic DNA was prepared based on a rapid small scale method (Dyall-Smith *et al.*, 1995). It was designed to produce DNA of a suitable quantity and quality for Southern blotting, but not for yielding high molecular weight DNA suitable for partial enzyme digests due to the small size of culture being manipulated. Owing to the osmotic properties of halophilic cells, this method of extraction is relatively simple to scale up, avoiding the common problems involved with the use of lysozyme to release the cell contents.

When analysed by gel-electrophoresis, the gDNA ran as a tight band equivalent in M_r to uncut λ DNA, implying it was of a suitable size and grade for use in restriction digests (Fig 3.1). When restricted with endonucleases, the DNA cut easily to give a distinctive and reproducible banding pattern and gave a 260nm/280nm ratio of 1.7. The DNA was resuspended to a concentration of 2mg/ml.

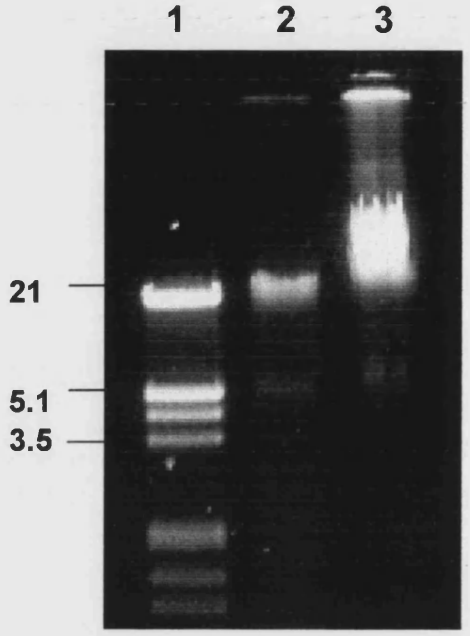


Figure 3.1
Analysis and Quantitation of Hf volcanii Genomic DNA.

Lane 1: λHindIII-EcoRI
Lane 2: *Hf volcanii* gDNA ~100 ng
Lane 3: *Hf volcanii* gDNA ~ 2μg

Fragment sizes are indicated in kilobases.
DNA was electrophoresed on a 1% (w/v) agarose gel.

3.3 Southern Analysis

3.3.1 Preparation of Genomic Hf volcanii DNA for Southern Blotting

A series of single and double restriction digests were carried out on *Hf volcanii* gDNA with a number of restriction endonucleases. The particular enzymes were chosen based upon the restriction map already obtained from the cloned *Hf volcanii* citrate synthase fragment. Both the single and double digests were run on the same gel in duplicate, with one half of the gel being blotted onto Hybond N^+ nylon membrane and being hybridised with a random α-^{32}P labelled 880 bp gene specific probe generated by PCR. Figures 3.2 and 3.3 show that the enzymes in the majority of the digests produced suitably sized DNA fragments on which to perform Southern analysis.

3.3.2 Preparation of an 880 bp Gene Specific Probe

The entire 880 bp cloned fragment of citrate synthase was to be used as a gene probe. A 10 ml overnight culture of *E.coli* XR1 Blue cells hosting the recombinant plasmid containing the citrate synthase gene fragment was used to sub-culture 100ml of LB and grown overnight. The plasmid DNA was extracted using a Qiagen plasmid 'maxi' prep system. Aliquots of 5μg plasmid DNA were digested with *EcoRI* and *HindIII* to remove the cloned citrate synthase fragment. The 880 bp gene fragment was gel purified and cleaned using the Bio 101 geneclean system.

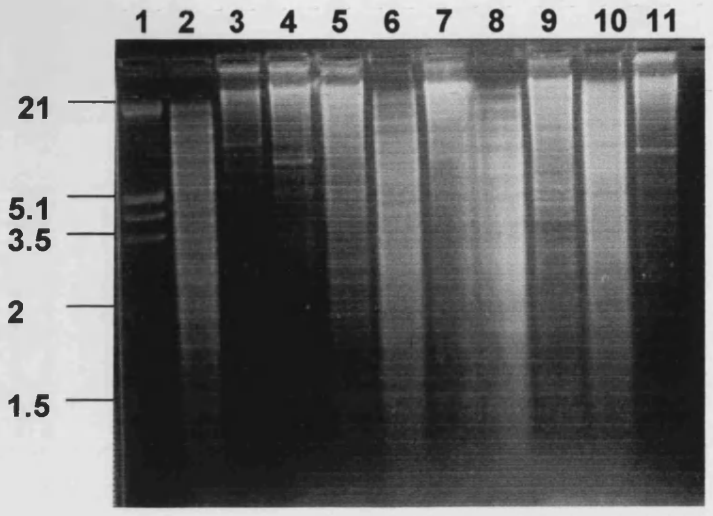


Figure 3.2
Genomic DNA Digested with Various Enzymes in Preparation for Southern Analysis.

Lane 1: λ/Hind/Eco marker
Lane 2: MluI
Lane 3: KpnI
Lane 4: HindIII
Lane 5: EcoRI
Lane 6: MiuI/KpnI
Lane 7: EcoRI/HindIII
Lane 8: EcoRI/MluI
Lane 9: EcoRI/KpnI
Lane 10: HindIII/MluI
Lane 11: HindIII/KpnI

Fragment sizes are indicated in kilobases.
The digests were electrophoresed on a 1% (w/v) agarose gel.
The gel was run at 15v overnight.

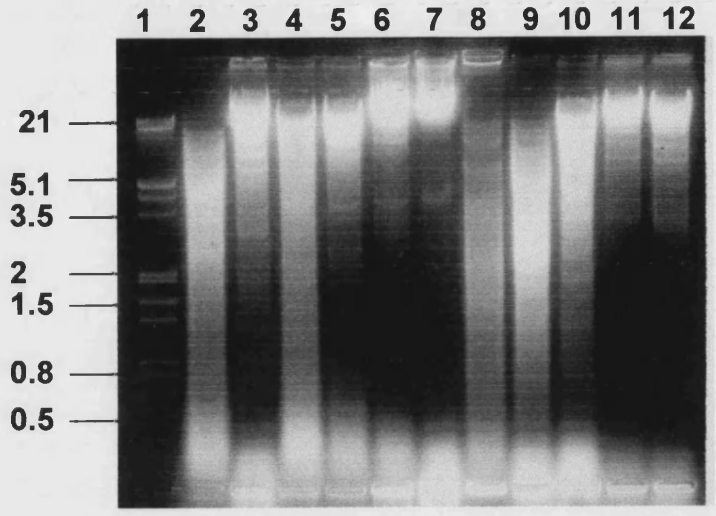


Figure 3.3
Genomic DNA Digested with Various Enzymes in Preparation for Southern Analysis.

Lane 1: λ/Hind/Eco marker
Lane 2: MluI
Lane 3: Pst1
Lane 4: Sau3AI
Lane 5: SphI
Lane 6: BamHI
Lane 7: SpeI
Lane 8: BssHII
Lane 9: BssHII/PstI
Lane 10: SphI/PstI
Lane 11: SpeI/PstI
Lane 12: BamHI/PstI

Fragment sizes are indicated in kilobases.
The digests were electrophoresed on a 1% (w/v) agarose gel.
The gel was run at 15v overnight.

3.3.3 Screening of the Southern Blots

Two Southern blots were hybridised with the 880bp fragment. The probe was labelled with α-^{32}P by the random prime method as described in Materials and Methods. Using the formula Tm = 4(G+C) + 2(A+T) the melting point (T_m) of the oligonucleotide is estimated to be in excess of 100 °C.

Both Southern blots were hybridised at 65°C for 16 hrs. The blots were washed with 1 X SSC containing 0.5% (w/v) SDS and 0.5 X SSC containing 0.5% SDS at 65°C. The corresponding autoradiographs are shown in Figures 3.4 and 3.5. A summary of the results is shown in Table 3.1.

Table 3.1 Summary of Deducible Results from Southern Blotting

(Taken from Figures 3.4 and 3.5)

Restriction Digest	Fragment size (kb)			
MluI	2	0.9		
MluI/KpnI	2	0.9		
EcoRI/MluI	3.5	2	0.9	
HindIII/MluI	4.5	3.5	2	0.9
BssHII	1.4			

* All other enzymes shown in Figures 3.4 and 3.5 produced autoradiograph fragments of 20 kb+

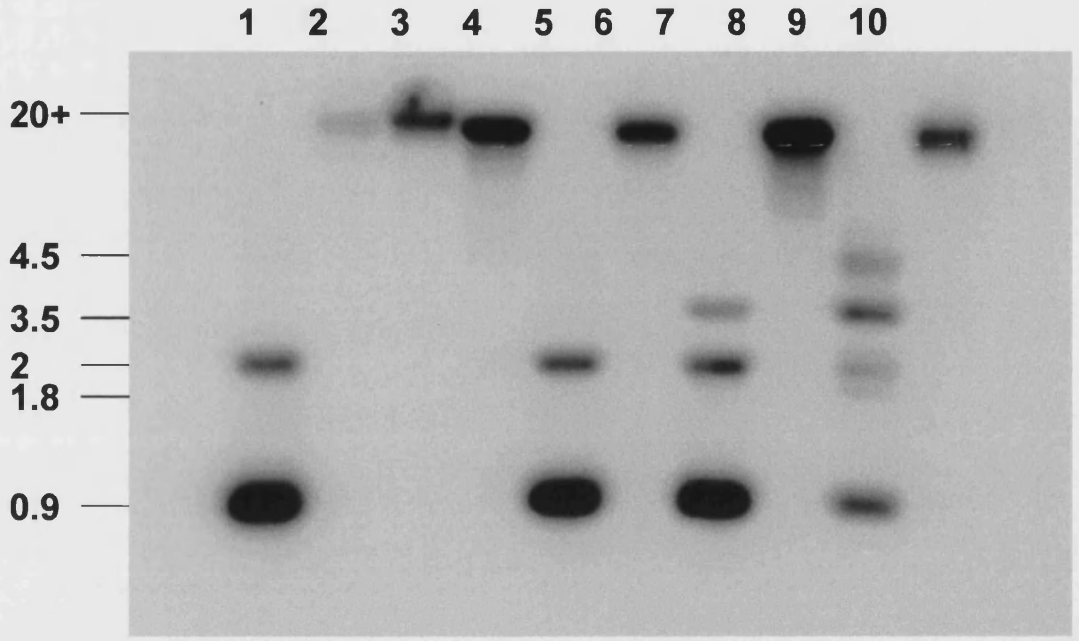


Figure 3.4
Autoradiograph of Southern Blot of Hf volcanii Restriction Digested gDNA Probed with the 880 bp Cloned Citrate Synthase Fragment.

Lane 1: MluI
Lane 2: KpnI
Lane 3: HindIII
Lane 4: EcoRI
Lane 5: MluI/KpnI
Lane 6: EcoRI/HindIII
Lane 7: EcoRI/MluI
Lane 8: EcoRI/KpnI
Lane 9: HindIII/MluI
Lane 10: HindIII/KpnI

Fragment sizes are indicated in kilobases.
The autoradiograph was exposed for 5 hours at -70°C.
The corresponding gDNA digests were run on a 1% agarose gel.

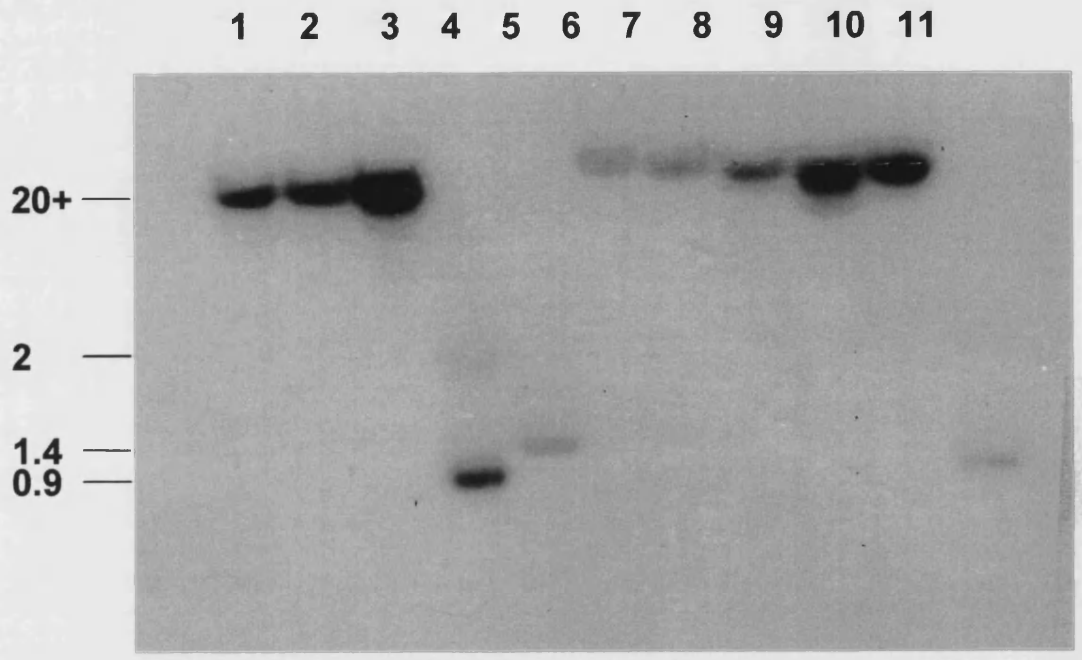


Figure 3.5
Autoradiograph of Southern Blot of Hf volcanii Restriction Digested gDNA Probed with the 880 bp Cloned Citrate Synthase Fragment.

Lane 1:Sau3AI
Lane 2: PstI
Lane 3: SphI
Lane 4: MluI
Lane 5: BssHII
Lane 6: SpeI
Lane 7: BamHI
Lane 8: PstI/SphI
Lane 9: PstI/SpeI
Lane 10: PstI/BamHI
Lane 11: PstI/BssHII

Fragment sizes are indicated in kilobases.
The autoradiograph was exposed for 5 hours at -70°C.
The corresponding gDNA digests were run on a 1% agarose gel.

As seen, many of the restriction enzymes created fragments too large to clone into a plasmid vector, with the majority of the probe hybridising to high M_r fragments or undigested DNA. However, clonable fragments were achieved by restricting gDNA with *MluI*, producing 2 clonable fragments of 0.9 and 2 kb and BssHII producing a single fragment of 1.4kb.

When gDNA was digested with *MluI*, the majority of the probe hybridised to the 0.9 kb fragment, consistent with the restriction map already obtained from the cloned CS fragment. When *MluI* was used in a double digest with *EcoRI* or *HindIII* other larger fragments appeared. These findings contradicted the results from the single *MluI* digest in conjunction with the restriction map. The significance of these results will be discussed later in the chapter.

3.4 Cloning of the 1.4 kb BssHII Fragment

The cloning of the 1.4 kb *BssHII* hybridising fragment (Fig 3.5) was attempted . It was removed from the corresponding duplicate gel and the DNA recovered using Bio 101 geneclean system. A commercially prepared plasmid, pBCSK+ (Stratagene) with a chloroamphenicol resistance marker, blue/white selection and a *BssHII* restriction site, was digested with the appropriate enzyme. The enzyme was heat denatured and the remaining plasmid prep treated with shrimp alkaline phosphatase. The fragment was ligated with the vector overnight and then transformed into *E.coli* XL1 blue, which was plated on to LB/Amp plates. No recombinant clones were detected. Subsequent digests of gDNA with *BssHII* and further cloning attempts of a 1.4 kb fragment all proved to be futile.

Therefore a new strategy was applied, namely the creation of a genomic library in λ phage. This method had successfully been employed in our laboratory with other Archaea and has the potential to produce a single clone containing the entire CS gene.

3.5 Preparation of a Hf volcanii genomic library in λ EMBL3

3.5.1 Digestion of Hf volcanii DNA for use in a genomic library

Approximately 10 ug of gDNA was digested with various concentrations of *Sau3AI* under controlled conditions at 25°C for 30 min. The DNA was restricted to suitably sized fragments using 10, 5, 2, 5, 1.25 and 0.63 units of the enzyme. Digests using lower concentrations of the enzyme resulted in incomplete digestion of the DNA. The results are shown in Figure 3.6.

Using the aforementioned concentrations of *Sau3AI*, scale-up digests were performed each using approximately 50 ug DNA (i.e. scaled-up by a factor of 5). Complete digestion was only achieved using 10, 5 and 2.5 units of the enzyme, probably due to the DNA being only partially dissolved. The results are shown in Figure 3.7.

3.5.2 Size fractionation of the digested DNA

DNA from the 3 scale-up digests in which complete digestion was achieved, were pooled, ethanol precipitated and resuspended in 500 μl distilled water. The DNA was size fractionated by passing through a NaCl gradient ranging from 5 – 10% (w/v) for 16 h. Fractions (500μl) were collected and run on a 1% (w/v) agarose gel (Fig 3.8). Fractions ranging in size between 7 and 20 kb, (lanes 2-7 Fig 3.8) corresponding to fractions 1-11, were selected and pooled. Combined fractions were diluted 1:2 with distilled water to minimize salt precipitation prior to ethanol precipitation and resuspension in 200 μl of distilled water. Combined size selected fractions are shown in Fig 3.9.

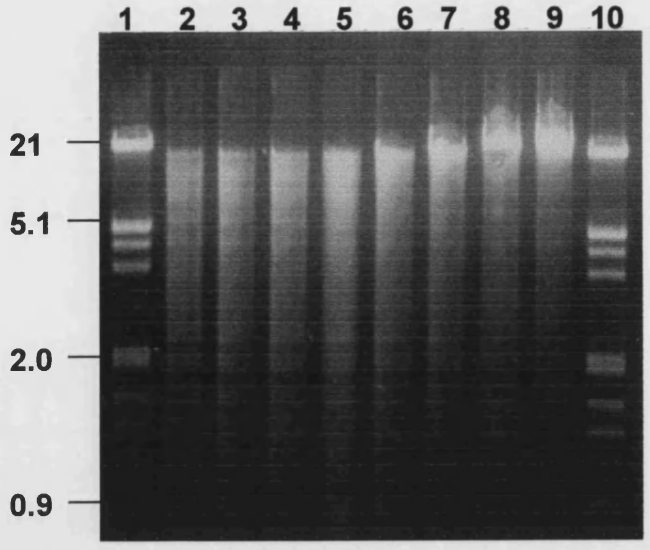


Figure 3.6
Preparation of Hf volcanii gDNA for a λ EMBL3 Library. Small Scale Partial Sau3AI Digests.

Lane 1: λHindIII/EcoRI
Lane 2:10 units
Lane 3: 5 units
Lane 4: 2.5 units
Lane 5: 1.25 units
Lane 6: 0.625 units
Lane 7: 0.313 units
Lane 8: 0.156 units
Lane 9: 0.078 units
Lane 10: λHindIII/EcoRI

Digests were electrophoresed on a 1% (w/v) agarose gel. Fragment sizes are indicated in kilobases.

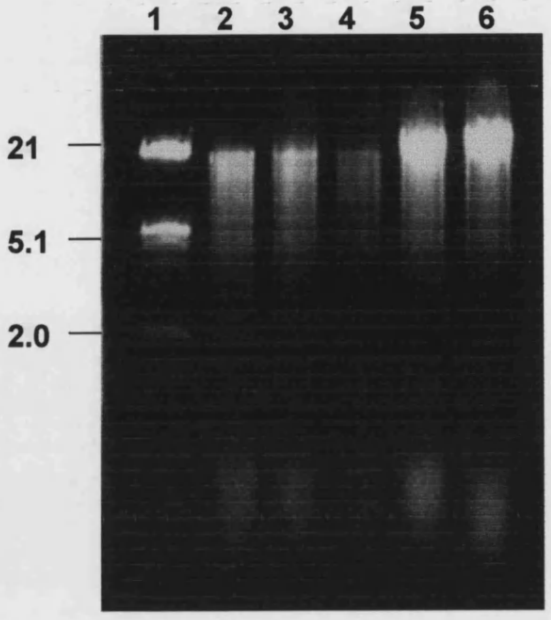


Figure 3.7
A 5X Scale up Sau3AI Partial Digest of Hf volcanii gDNA (Figure 3.6 lanes 2 - 6).

Lane 1: λHindIII/EcoRI
Lane 2:10 units
Lane 3: 5 units
Lane 4: 2.5 units
Lane 5: 1.25 units
Lane 6: 0.0625 units

Digests were electrophoresed on a 1% (w/v) agarose gel. Fragment sizes are indicated in kilobases.

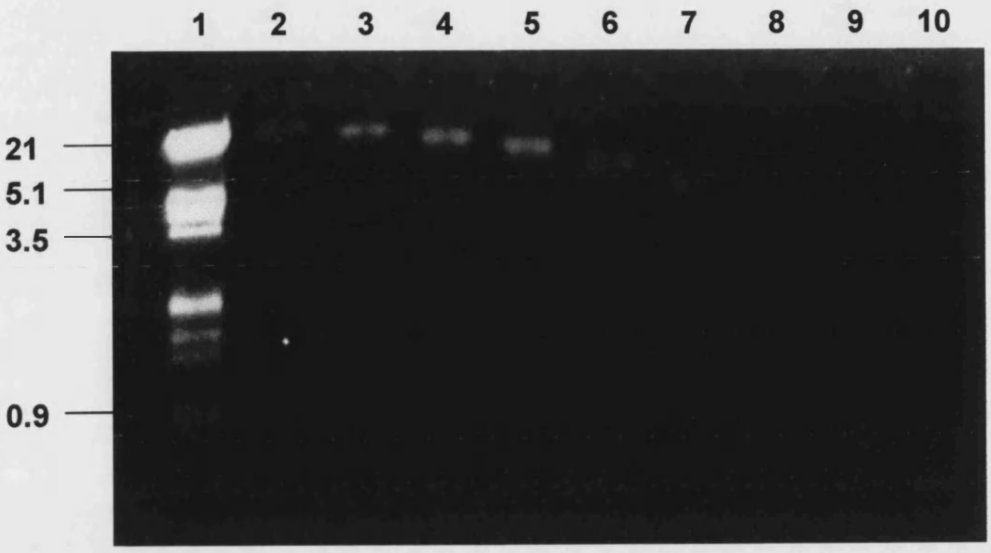


Figure 3.8
Size Selected Fractions of Hf volcanii gDNA After Centrifugation Through a NaCl Gradient.

Lane 1: λ HindIII/EcoRI
Lane 2: fraction 1 (500 μl)
Lane 3: fraction 3 (1.5 ml)
Lane 4: fraction 5 (2.5 ml)
Lane 5: fraction 7 (3.5 ml)
Lane 6: fraction 9 (4.5 ml)
Lane 7: fraction 11 (5.5 ml)
Lane 8: fraction 13 (6.5 ml)
Lane 9: fraction 15 (7.5 ml)
Lane 10: fraction 17 (8.5 ml)

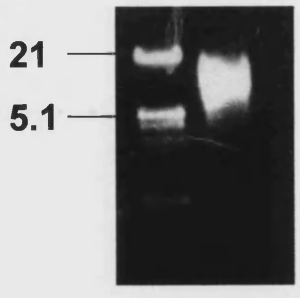


Figure 3.9
Combined fragments (Fig 3.4 lanes 2-7) corresponding to fractions 1-11 of partially digested Hf volcanii gDNA.

Fractions were electrophoresed on 1% (w/v) agarose gels. Fragment sizes are indicated in kilobases.

3.5.3 Packaging of the phage heads

Size selected DNA (1μl) was ligated into λ EMBL3 arms, prepared by digestion with HindIII, and packaged into phage heads using Gigapack® III gold packaging extract (Stratagene). Chloroform was used to stop the packaging reaction and to preserve the λ stock solution. Packaged phage heads were used to infect host cells *E.coli* X-LI blue MRA (P2) in the presence of 10 mM $MgSO_4$ and 0.2% (w/v) maltose. Infected host cells were then combined with top agar and plated out on LB/agar plates. The library was titrated in 10X serial dilutions of the packaged phage heads; the packaging efficiency was found to be 2 $x10^6$ pfu/ml.

3.5.4 Screening of the library

Plates were made using a 1 in 10 and a 1 in 100 dilution of the packaged phage heads. 10 plates were made of each dilution, plates numbered 1 to 10 corresponding to 1 in 10 dilutions of the packaged phage heads, and 11 to 20 to 1 in 100 dilutions of the library. Plaque DNA was transferred to hybond N^+ nylon membranes and the library screened with the 880bp cloned CS fragment labelled with α-^{32}P as previously used in Southern analysis.

Hybridisation was carried out at 65°C for 16 h. Membranes were washed with 1 X SSC with 0.5% (w/v) SDS and 0.5 X SSC with 0.5% SDS at 65°C. Two X-ray films were exposed to the membranes at –70°C ; one was developed after 5 h exposure and the second after 36 h. The resulting positive autoradiographs are shown in Figure 3.10. Positive plaques were selected from the original plates and used as the basis for a secondary screen. Twenty plaques were taken in total and re-screened with the labelled probe; those giving positive signals are shown in Figure 3.11.

Plaques were taken from plate 9 as all plaques present gave a strong signal on the respective autoradiograph. Phage stocks were also taken from the other plates shown in fig 3.11 in the event of a double insert or partial gene fragment being found in clone 9.

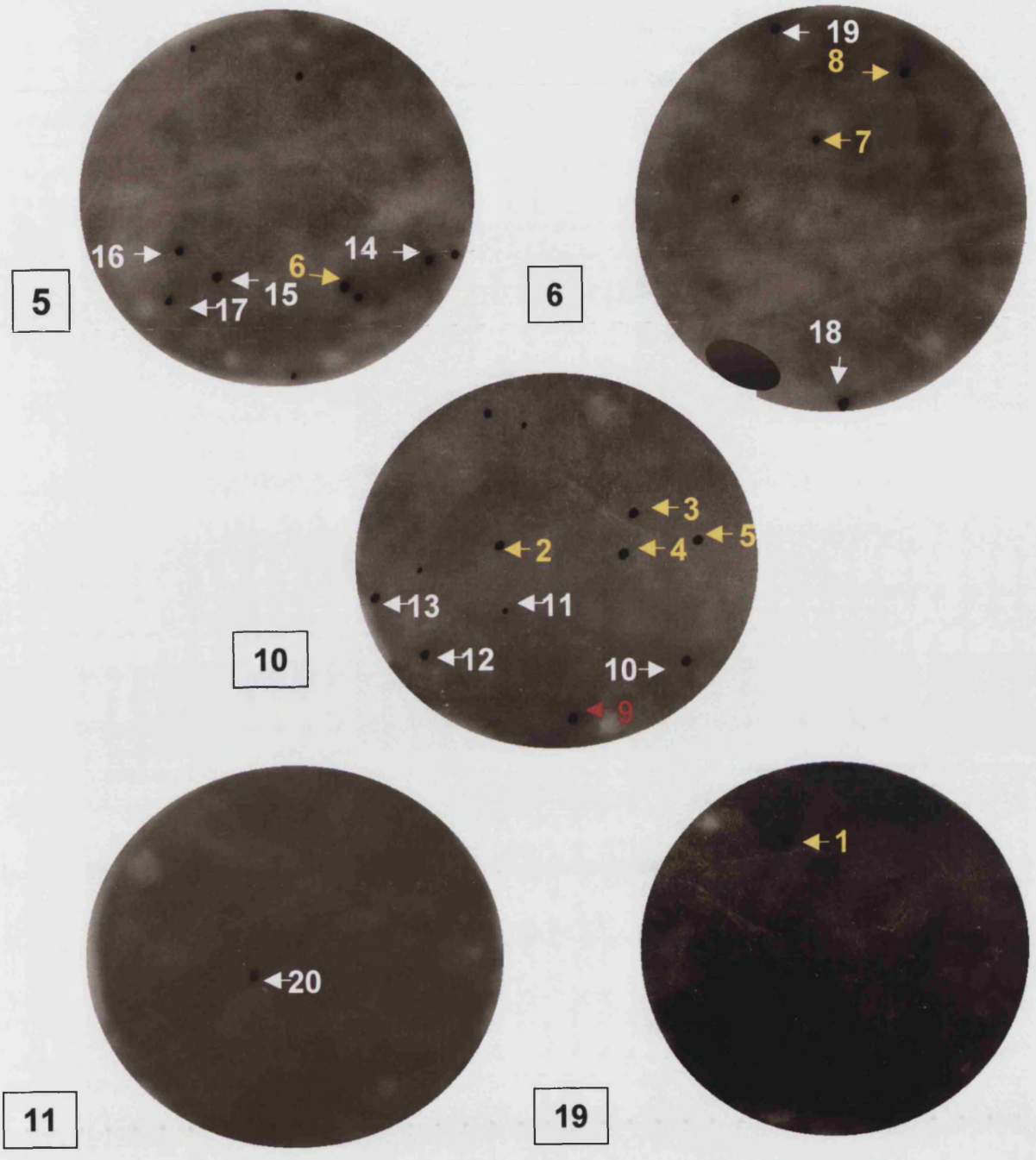


Figure 3.10
Autoradiographs from Primary Screening of Hf volcanii Genomic Library in λ EMBL3.

Boxed numbers correspond to original plate numbers.
Plaques labelled yelow were visible on the autoradiograph after 4 hr exposure at -70°C.
Plaques labelled white were visible on the autoradiograph after 60 hr exposure at -70°C.
Red plaque is the plaque from which the complete sequence was obtained.

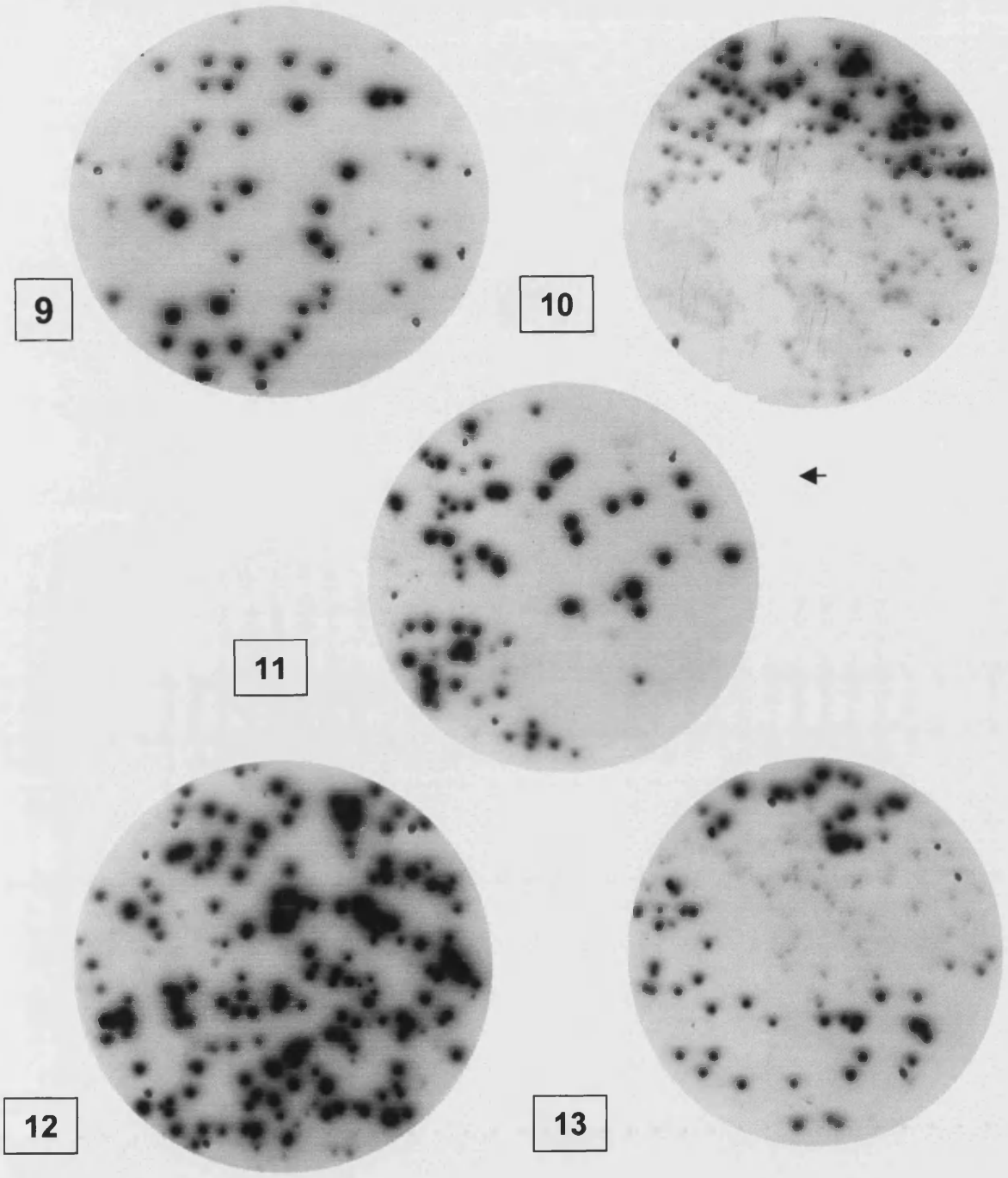


Figure 3.11
Autoradiographs from Secondary Screening of Hf volcanii Genomic Library in λ EMBL3.

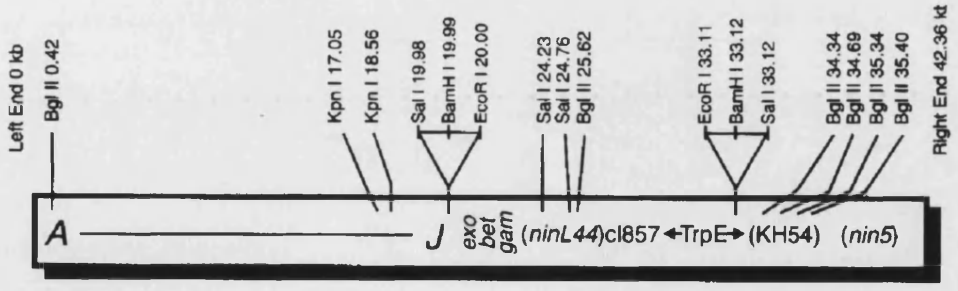
Boxed numbers correspond to selected plaque numbers (see Fig 3.10).
X-ray film was exposed to the membranes for 24 hours at -70°C.
Plaques were taken from plate No. 9 and used to prepare λ DNA.

3.5.5 Preparation of λ DNA

Lambda DNA was made from clone 9 of the secondary screen using 300 ml of cell lysate and a Qiagen λ maxi kit. Extra RNase and DNase were added to the cell culture after initial infection of the *E.coli* host cells to ensure complete removal of all contaminant host cell nucleic acid, which could prove to be problematic when sequencing. Chloroform was used to ensure complete cell lysis. Resulting λ DNA was run on a 1% agarose gel against a λ marker to gauge quantity and quality of the sample. (Fig 3.12)

3.5.6 Sizing of the λ clone insert

Lambda clone 9 (1μg) was digested overnight at 37°C with restriction endonucleases. The enzymes were chosen based on the restriction map of the λEMBL3 arms (Fig 3.13). Digests were performed overnight at 37°C. DNA fragments can be seen (Fig 3.14) and a summary of the results are presented in Table 3.2. As most of the digests show , the largest fragment produced was larger than the highest mw marker, making sizing of the insert difficult. Digestion with *KpnI/BglII* produced fragments below 21 kb and the size of the insert was estimated to be approximately 15 kb.



reproduced from the Stratagene catalogue

Figure 3.13 Restriction map of λ EMBL3 arms

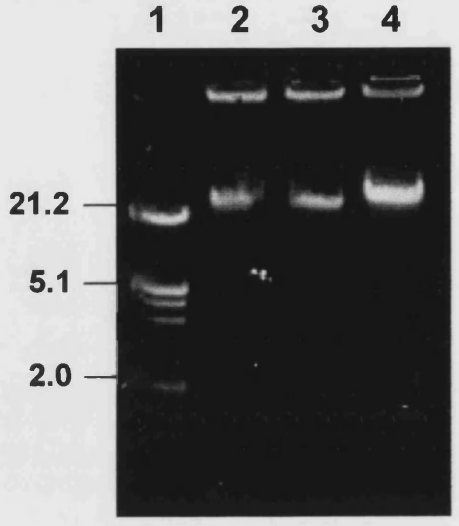


Figure 3.12
Analysis and Quantitation of λ Clone DNA Prepared using Qiagen® 500 Maxi Preps.

Lane 1: λ/H/E marker
Lane 2: λ clone 9 DNA prep 1 ~250 ng
Lane 3: λ clone 9 DNA prep 2 ~250 ng
Lane 4: λ clone 9 DNA prep 3 ~500 ng

From the appearance on the gel , the DNA preparation in Lane 4 seems to be of the higher quality and quantity/μl.

The DNA was electrophoresed on a 1% (w/v) agarose gel. Fragment sizes are indicated in kilobases.

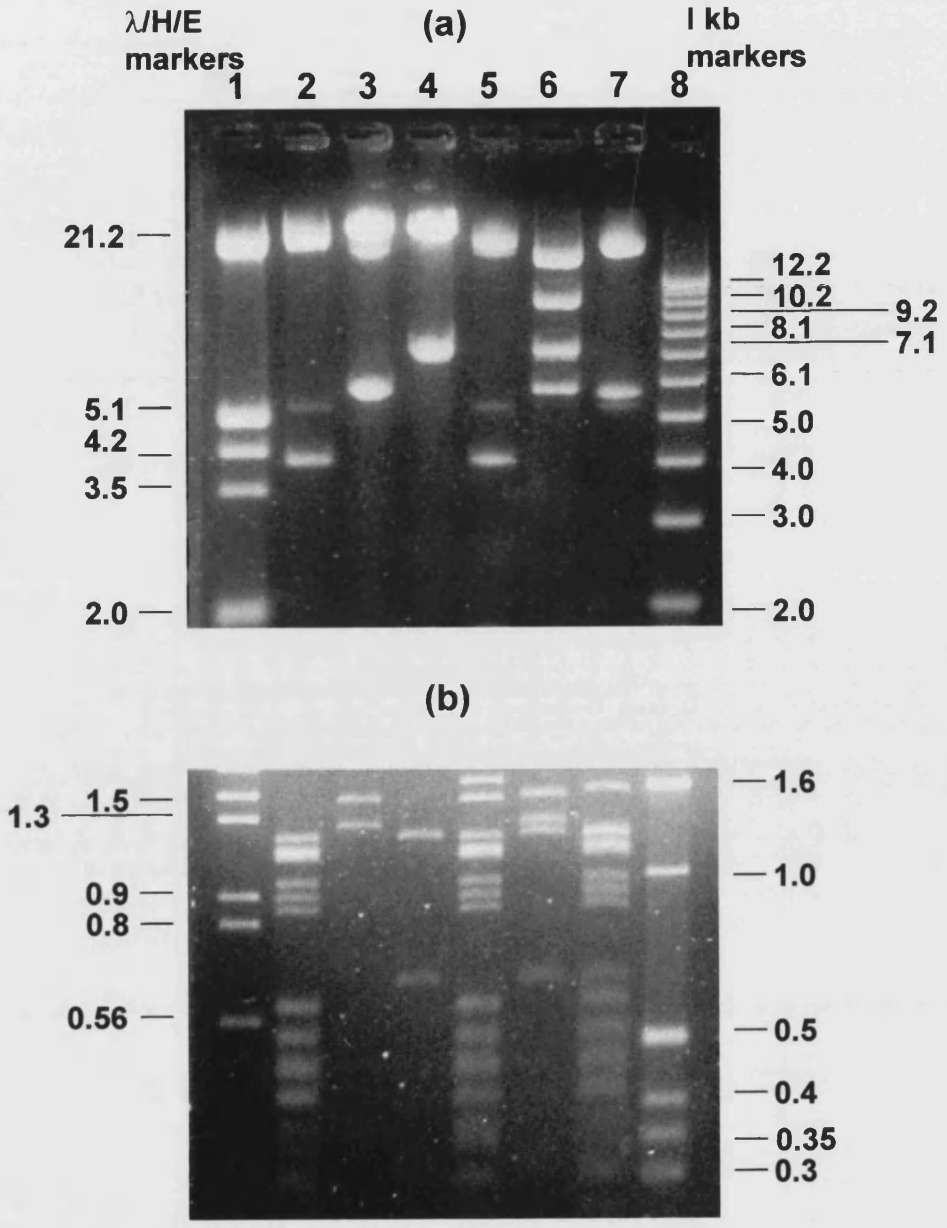


Figure 3.14
Sizing of the λEMBL3 Clone 9 Insert.

Key: λ clone 9 was digested with various enzymes

Lane 1: λ/H/E marker
Lane 2: XbaI
Lane 3: XhoI
Lane 4: Bgl II
Lane 5: BamHI
Lane 6: KpnI
Lane 7: Sal I
Lane 8: 1 kb marker

Fractions were elctrophoresed on 0.8% (w/v) (a) and 2% (w/v) (b) agarose gels.
Fragment sizes are indicated in kilobases.

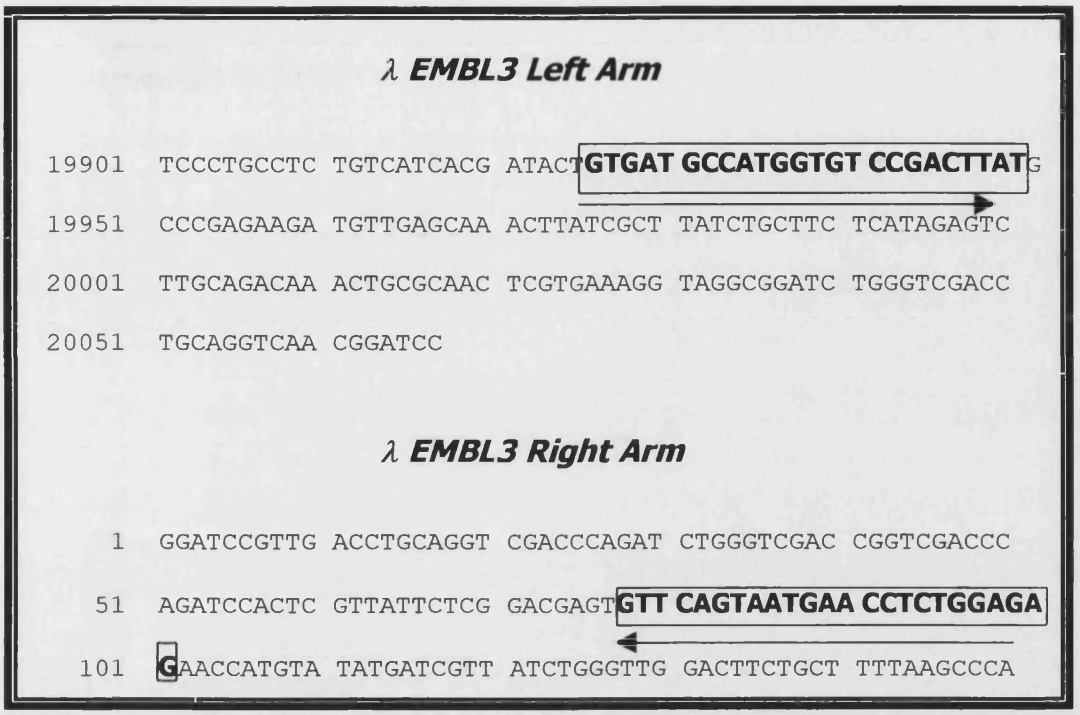
Table 3.2 Results from Digestion of λ Clone 9 to Size the Genomic Insert.

Restriction Enzymes used					
EcoRI	*BamHI*	*Bgl II*	*Eco/Bam*	*Bgl /KpnI*	*Eco/Bgl*
Fragment Sizes (kb)					
21+	21+	21+	21+	17	21+
5.2	20	7	5.2	10.5	5
4.1	5.5	1.2	4.1	7	1.2
1.2	1.5	0.6	1.7	5	1.15
1.15	1.3	0.32	1.5	1.5	1.1
1.1			1.2	1.3	0.95
0.95			1.15	1.2	0.9
0.9			1.1	0.6	0.85
0.85			0.95	0.32	0.6
0.6			0.9		0.54
0.54			0.85		0.45
0.45			0.6		0.4
0.4			0.54		0.38
0.38			0.45		0.34
0.34			0.4		0.3
0.3			0.38		
			0.34		
			0.3		

3.6 Sequencing of the λ Clone

40 kb λ DNA of a suitably high grade is capable of being directly sequenced. Primers (EMBL3L and EMBL3R) based on λ EMBL3 arm sequence were designed to gauge whether the quality of λ DNA produced was suitable for sequencing (Fig 3.15). Sequencing using these primers was successful, not only proving that the prepared clone9 λDNA was of a suitable quality to yield sequencing results, but that the gene of interest was not truncated at the vector/insert junction.

Figure 3.15 Primers based on λ EMBL3 sequence



Binding positions of primers EMBL3L and EMBL3R shown on the respective λ arms. Numbers correspond to the positions within the arms. The left arm is 20,068 bp and the right arm 9,170 bp in length. Sequence was taken from EMBL/GenBank database.

3.6.1 Sequencing the complete gene

Additional sequencing with gene-specific primers also proved successful, not only completing the citrate synthase gene sequence but also in providing sequence data for the N-terminus for which previously only the amino acid sequence was known. With the culmination of the sequenced PCR product and the sequence data from the λ clone, the gene sequence was now complete. However, it was necessary to have the complete gene sequence from a genomic clone rather than partially from the PCR fragment. Therefore two primers were designed to provide the complete double stranded sequence of the citrate synthase gene from λ clone 9. Four primers were used to complete the sequencing of the gene and a further two to investigate up and downstream of the coding region. The primer sequences are listed in Chpt 2 Table 2.11 and their binding positions within the complete nucleotide sequence are shown in Fig 3.16. The complete nucleotide sequence in conjunction with the designated promoter region and coding amino acid sequence will be discussed further and can be seen in Chpt 4.

Figure 3.16 The complete citrate synthase sequence

Sequences were aligned and edited using the GCG package. The full sequence of the citrate synthase gene with up and downstream regions is shown.

```
        -415  TGGCGTGGAG GACGGCGGAG TACGAGAGGT CGATGAGTAA CTCCGCGGTG
                                                      <-----
        -365  TCTTTCATCT CGGCGAGGAC GGCCTTGACG CTGACAGGCT CGTACTCGAT DM7
        -315  GTCCTCCGGA TTCATGCGCC CGCCTTCACG ACGGGACGGC AAAAGGCTTG
                                  ------------------------>
        -265  CCAGGAGTAA CGGTGCCACC GTGGTTCGAG AGAACTACGC CGTCGGGGGA DM2
        -215  TATGAGTTTG CAGGTCGCCC GAGCGGCGCG CGCGCGACCC ATCGGGGGGA
        -165  CTCCGTATAA GTGCCCCGAA AATAACGACA ATCGTCGAAC ACTATTTTAG
        -115  ACACCGACTA TCCGCGAAAC CGACAGACGG GGCTGCCACT GTATGTGTAT
         -65  GGCGGTCGGT AACAGATGAG TGAAAACGAT AACGCTTTTT TGATATGGTA
         -15  GGCAACGGTC TGGGTATGTC AGGCGAACTG AAGCGGGGGC TGGAAGGTGT
          36  GCTGGTCGCC GAATCGAAAC TCAGTTTCAT CGACGGCGAC GCGGGACAAC
          86  TCGTTTACTG CGGGTACGAT ATCGAAGACC TCGCACGGGA CGCAAGTTAC
COSAS    136  GAGGA<-----AGTAC TCTATCTCCT GTGGCACGGA GCGCTTCCGA CGGGCGAGGA
         186  ACTCGACGCG TTCTCCGACG AGCTCGCGGC CCACCGCGAC CTCGACGACG
                                  ------------------------>
         236  GCGTCCTCGA CGTGGCACGC GAACTCGCCG AACAGGACGA GTCGCCGATG DM8
         286  GCGGCGCTCC GAACGCTCGT CTCGGCGATG TCGGCGTACG ACGAAAGCGC
         336  CGACTTCGAG GACGTGACCG ACCGCGAGGT CAACCTCGAG AAGGCAAAGC
         386  GCATCACGGC GAAGATGCCG TCGGTGCTCG CGGCCTACGC CCGCTTCCGT
         436  CGCGGCGACG ACTACGTCGA ACCCGACGAG AGCCTGAATC ACGCGGCGAA
         486  CTTCCTCTAC ATGCTCAACG GCGAGGAGCC GAACGAGGTG CTCGCCGAGA
         536  CGTTCGACAT GGCGCTCGTG CTCCACGCCG ACCACGGACT GAACGCCTCG
         586  ACCTTCTCCG CGATGGTCAC GTCCTCGACG CTCTCTGACC TCTACAGCGC
         636  AGTCACGTCC GCAATCGGCA CGCTCTCGGG GTCGCTCCAC GGCGGCGCGA
         686  ACGCGAACGT CATGCGGATG CTGAAGGACG TCGACGACAG CGACATGGAC
         736  CCGACCGAGT GGGTCAAAGA CGCCCTCGAC CGCGGCGAGC GCGTCGCGGG
         786  GTTCGTCCAC CGCGTCTACA ACGTCAAGGA CCCCCGCGCG AAGATTCTCG
         836  GCGCGAAGGC CGAGGCGCTC GGCGAGGCCG CCGGCGACAT GAAGTGGTAC
         886  GAGATGTCGG TCGCCATCGA GGAGTACATC GGCGAGGAGA AGGGCCTCGC COSS
                                                      <-----
         936  GCCGAACGTG GACTTCTACT CCGCGTCGAC GTACTACCAG ATGGGCATCC DM9
                                  <------------------------
                                                      ------>
         986  CCATCGACCT CTACACTCCC ATCTTCGCCG TCTCGCGCGC CGGCGGCTGG
```

```
        1036  ATTGCGCACG TCCTCGAACA GTACGAGGAC AACCGCCTCA TCCGCCCCCG
        1086  CGCCCGCTAC ACCGGCGAGA AAGACCTCGA CTTCACGCCG GTCGACGAGC
        1136  GATAGGCGAC GAACCGCACG CGACTGTTTT CCGTCTGCTG CCCCCGGAGC
         +48  GACGCGTCGT TCCCTTCACC CTCCGCCTCC CGCTCTCGCG TTCGCCGTCC
         +98  GCCGCACGCG CCGCCCTCAG TCGTCTTCGC CCACGTCGTC GACGCCGACG
        +148  CTCCCGCGTT CGACCATGAG CGAGTCGGCG TGGCCCGCCC AGTCGGCGTA
        +198  GGAGAACGCG ACGAGCGCGA ACAGCCCCAT CCACGTGAAC GACAGGACGA DM3
DM4     +248  TGCCGATTCG CGCGCCGAGC GGGCCGTAGC CGAACCCCTC GACGAGCAGG
        +298  TACGACAGGC CGACGAAGAA GCCGAACACC CCGGTCAGCC GGGCGACGAA
        +348  GGGAATCCGC GTCGCGCCCG CGCCCTGCAA CGACCCGGAG AGGCCGACGA
        +398  ACAGCGCCAG CCCCGGTGCG GCGAGACCGT AGGCGACGGC GAACTTCACG
        +448  GCCCAGTCGA GGTCCGTGCC GCCGCCGGAG CCGACGAGGC CGACGATGGG
        +498  TTCGGTGAAC GCCGCGAGCA GGAGCCCGAT GCCGCCGACC GTGACGACCG
        +548  AAAGCAGCGC CACCGCGCCG CCCTGATACC GCGCGGTTTC GCTGTCGCCC
        +598  GCGCCGAGCG CCTGCCCGAC GAGGACGCTG GCGGCGACGT TGTAGCCGCG
        +648  CGCCAGCGGG CTTGCGACCT GCTGGTAGAC CCGGCGGCCG ATTTGGAAGC
        +698  CGGCGTTGAC GCGTCACCGA AGCCGAGGAG CAGGGCGTTG AACGGGAACT
        +748  CTGCGAGTTC GGCGGCGATA CCCTCGCCGA CCCGCGGGGT GGCGATACGG
        +798  ACAATTGGCC GGCGACCGTC CGGTCCCGGG GCGGACCACC AAGTCGGTGA
        +848  CCCGGGCGGC AAACGCGACA TAACCATCCG CCGAAAAGAC TTTTGCGTCA
        +898  AATTGCAC
```

The complete citrate synthase gene sequence. The primers are shown in bold with arrows to indicate the direction in which they are sequencing. The translation start site and the stop codon are highlighted within boxes. Red sequence indicates that which was originally obtained from the PCR product, and blue that which was exclusive to the λ clone. Sequence up and downstream of the gene is indicated by – and + signs respectively.

The sequence shown here is the finalised sequence of the citrate synthase gene, it has been sequenced on both nucleotide strands. This was not a straight forward process, the problems encountered during sequencing of the complete gene, and the method employed to gain the complete double-stranded sequence is discussed next.

3.6.2 Enhancement of Sequence

Direct sequencing of the gene within the 40 kb λ clone was not straightforward, especially as *Haloferax* DNA is GC rich. The DNA needed to be of a high quality and very clean for sequencing to occur. This was especially important when using high concentrations as was required here. Sequencing tended to be highly temperamental, and primers which had

shown to bind to the genomic templates in previous PCR reactions would not produce sequence. The DNA was known to be of sequencable quality because primers based on the λ arms had successfully given sequence data.

It was thought that this problem may be due to the formation of secondary structures due to the high (G+C) content of the *Haloferax* genome (Tindall *et al.*, 1984, Gutierrez *et al.*, 1989, 1990, Charlebois *et al.*, 1991). Attempts were made to overcome this problem by using 5% DMSO in sequencing reactions and boiling samples for 5 min prior to sequencing; this did improve some of the sequences but was not successful for all of them.

3.7 Subcloning Part of the Gene

The problems encountered in sequencing *Hf volcanii* DNA were overcome by subcloning part of the λ clone.

A restriction map was provided by the culmination of the primary sequence obtained from the λ clone and the original PCR product; two *BssHII* and *MluI* sites were identified flanking the problem area. These were previously highlighted in Southern analysis.

Digests of λ clone 9 DNA using *BssHII* (Fig 3.17) were conducted at 37°C overnight. Digested DNA was run on 1% (w/v) agarose gels, fragments of the appropriate sizes were removed and the DNA extracted using the Geneclean system.

DNA fragments were ligated in to a pUC based vector pBC SK+ (Fig 3.20) the resulting plasmid was transformed into *E.coli* XLI-blue. Utilising the blue/white colony selection protocol, potential recombinant clones were selected, grown overnight and plasmid DNA prepared using the SNAP™ miniprep system.

One recombinant clone, originating from the *BssHII* digest, contained an insert of the expected size, (Fig 3.18 lane 2). This and the other clones present in Fig 3.18 were digested with *BssHII* to see if they contained an insert (1.2 kb) and vector (3.4 kb) of the appropriate sizes (Fig 3.19). The larger of the three recombinant clones (Fig 3.18 lane 2) was seen to contain an insert and vector of the correct size (Fig 3.19 lane 4), authenticity of this clone was provided by sequencing. Double stranded sequencing of this region was made possible by using M13 forward and reverse universal primers and gene specific primer DM8, see Fig 3.21.

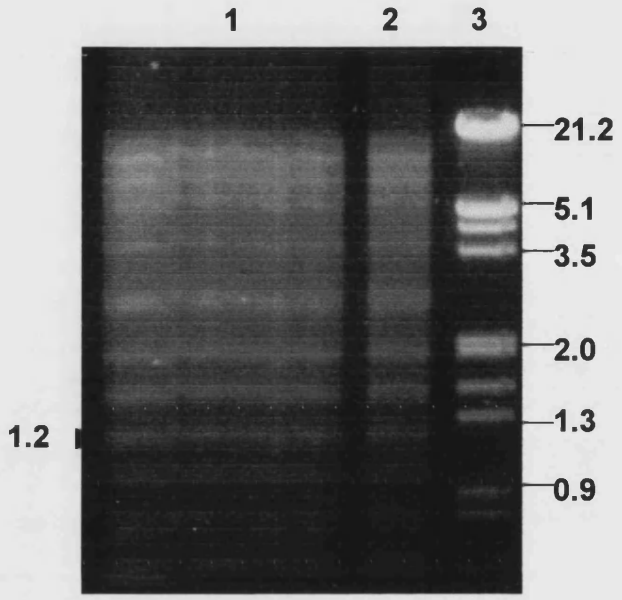


Figure 3.17
λ Clone 9 Digested with BssHII for Use in Subcloning into a Plasmid Vector.

Lane 1: Excess λ (9) digested with BssHII which was eventually cut out and cloned.
Lane 2: λ (9) digested with BssHII, used as a reference.
Lane 3: λ/H/E marker

Fragments were electrophoresed on a 1% (w/v) agarose gel.
Fragment sizes are indicated in kilobases.
Arrows indicate the fragment of interest.

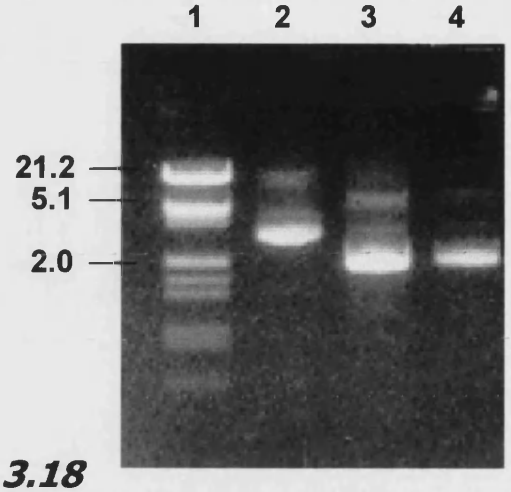


Figure 3.18
Undigested Recombinant Plasmid Subclones of λ Clone 9 (BssHII Digests).

Lane 1: λ/Hind/Eco marker
Lane 2: recombinant plasmid thought to contain insert
Lane 3+4: Recombinant plasmids thought to contain no inserts

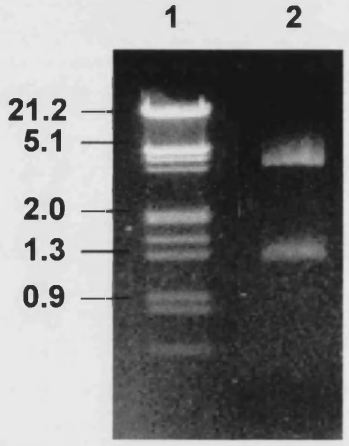
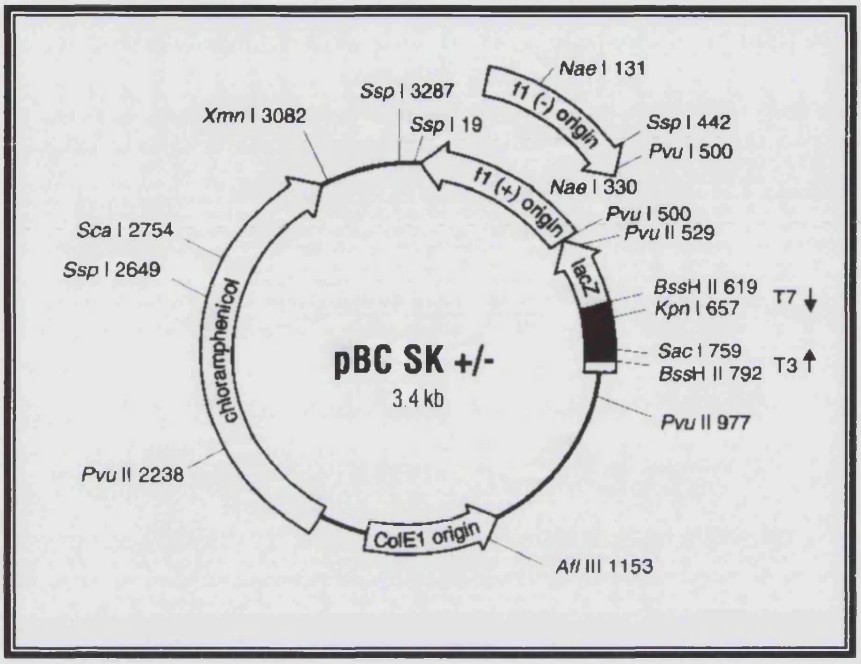


Figure 3.19
Recombinant Plasmid Subclones of λ Clone 9 Digested with BssHII.

Lane 1: λ/Hind/Eco marker
Lane 2: BssHII digested recombinant plasmid (Fig 3.18 lane 2) shown to contain a 3.4 kb vector and 1.2 kb insert.

Fragments are run on a 1% (w/v) agarose gel.
Fragment sizes are indicated in kilobases.

Figure 3.20 Plasmid vector pBC SK+



Reproduced from the *Stratagene* catalogue

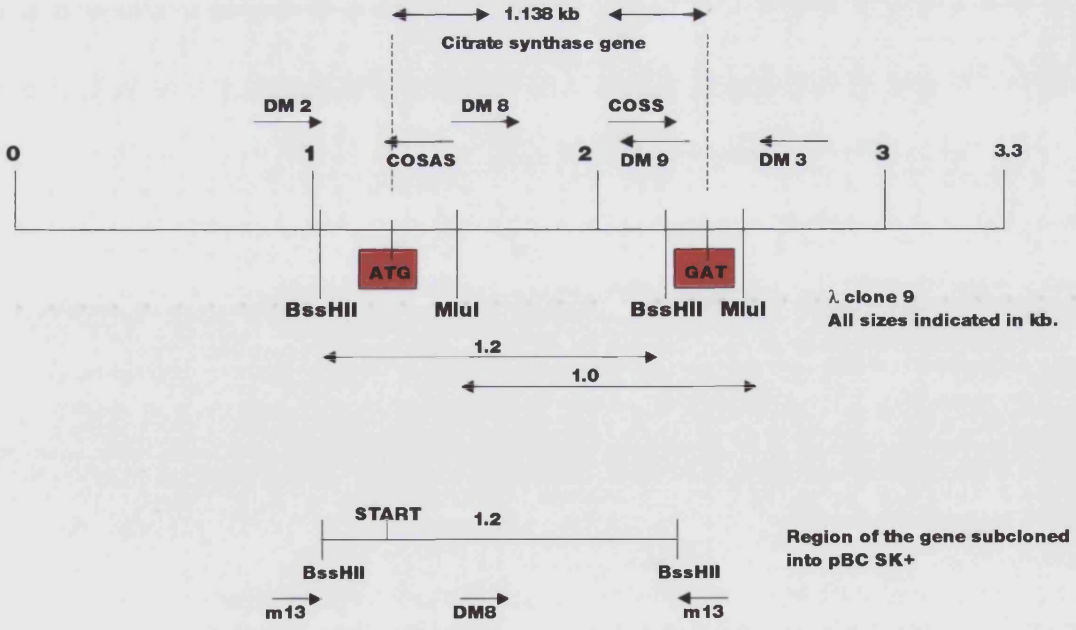


Figure 3.21 Sequencing Strategy of the Hf volcanii Citrate Synthase Clone

3.8 Discussion

This chapter describes the cloning and sequencing of the citrate synthase gene from *Hf. Volcanii.* Initial attempts to achieve this favoured the Southern blot approach. Although efforts to complete the gene sequence by this method eventually proved futile, hindsight in the light of the complete sequence divulged some interesting results.

It is now known that the citrate synthase gene contains 1 *MluI* and 1 *BssHII* site, and that there is a second occurrence of each site outside the coding region, in up and downstream sequences. These findings are in agreement with results from Southern analysis, and were eventually used to subclone part of the gene. Therefore, when digested with MluI, the complete gene is contained within two fragments of 2 kb and 1 kb, seen in Fig 3.4 with the majority of the sequence being contained within the 1 kb segment. *MluI* used in conjunction with *EcoRI* or *HindIII* should still provide the citrate synthase gene in two fragments as the restriction map of the gene does not indicate the presence of the latter two sites. However, the results (Fig 3.4 lanes 7 and 9), contradict this observation.

The type of oligonucleotide probe used in hybridisation of the Southern blot was long and non-degenerative. Specificity was attained not only in its length, but also in the fact that it was a replica of a large percentage of the gene. This gave all of the oligonucleotide molecules the potential to hybridise to the target, thus yielding a strong signal. Non-specific hybridisation was potentially counteracted by high hybridisation temperatures (65°C) and very stringent washes 1% and 0.5% (v/v) SDS, also conducted at 65°C.

The construction of the library went to plan with no major problems. The key to the successful construction of a library lies within the quality and quantity of the DNA. Prior to digestion *Hf. volcanii* DNA showed little evidence of shearing, with the DNA electrophoresing in a tight band equivalent in size to uncut λ. The DNA was available in abundance, making size selection after centrifugation through a NaCl gradient an easy procedure.

Preparation of λ clone DNA tended to be temperamental. Best results were achieved when DNA was prepared from larger volumes of lysate (optimum 300 ml). DNA could be extracted from smaller volumes of lysate but it tended to be of a lower concentration and sequence was never successfully obtained from it. This may be due to the volume required for sequencing; that is in order to obtain 750 ng in a 12 μl sample volume, the DNA had to be concentrated, thus magnifying the effect of any impurities present that might affect the sequence.

Results from digestion of λ clone 9 (Fig 3.14, Table 3.2) show the genomic insert to be approximately 15 kb. This is consistent with both the size originally selected and the packaging capability of the λ phage heads. We are confident that the gene itself is complete and is not a product of two fragments ligating together, as it translates into a continuous ORF of approximately the expected size ~1.3 kb based on other known CS gene sequences and contains no *Sau3AI* sites within the coding region.

DNA that appeared to be clean, non-sheared and of a suitable concentration did not always give satisfactory sequence. In many cases chromatogram quality was improved by boiling the sample in the presence of 5% DMSO prior to sequencing. Many of the problems encountered during sequencing were attributed to secondary structure formation. This is consistent with improved sequencing results after boiling with DMSO and the fact that the Haloferax genome has a high G+C content.

Certain sequencing primers (DM9) consistently gave poor quality chromatograms from λDNA, making it necessary to subclone part of the gene. Poor quality sequence was still obtained from the subcloned plasmid pBCSK+. The primer was seen to bind to the template as chromatograms were inevitably produced, as were PCR products of the expected size when used with a suitable sense primer.

The complete citrate synthase gene was 1.138 bp in length translating into a 380 amino acid protein. The G+C content of the gene was found to be 60%. Up and downstream regions containing potential control elements will be discussed in the next chapter along with protein structure and function.

Chapter 4

Analysis of the Haloferax volcanii Citrate Synthase Gene

4.1 DNA Sequence Analysis

The complete citrate synthase gene was found to be 1.138 kb in length, translating into a 380 amino acid protein. The *Haloferax* genome has previously been mapped and studied (Charlebois *et al.*, 1989; St. Jean & Charlebois, 1996) and the G + C content of halophilic genes is generally known to be high; citrate synthase proved to be no exception, the G + C content being found to be 60%. The gene itself was not only sequenced on both strands but up to 500 bp both up and downstream of the coding region were also sequenced to look for evidence of control elements and the presence of other genes.

4.1.1 Halobacterial Regulatory Regions

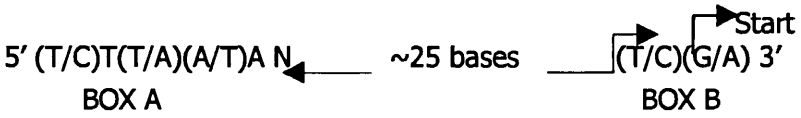
The transcriptional apparatus of the Archaea more closely resembles that of Eukarya than Bacteria (reviewed by Zillig *et al.*, 1993). Extensive characterization of the transcription machinery of the Archaea indicates that in terms of immunological cross-reactivity, primary sequence and subunit structure, their RNA polymerases more closely resemble the RNA polymerases II and III of Eukarya than the RNA polymerases of Bacteria (Langer *et al.*, 1995; Zillig *et al.*, 1989). Unlike the relatively simple bacterial RNA polymerases, the archaeal and eukaryal RNA polmerases display complex subunit compositions, with archaeal RNA polmerases typically containing 12 or more subunits. Archaea also possess homologues to the genes encoding eukaryal transcription factors TFIIB (Danner & Soppa, 1996, Creti *et al.*, 1993; Ouzounis & Sander, 1992) and the TATA box-binding protein TBP (Qureshe *et al.*, 1995; Marsh *et al.*, 1994; Rowlands *et al.*, 1994).

It has been proposed that that an AT-rich sequence with similarity to the eukaryal TATA box is important for promoter function in the Archaea (Reiter *et al.*, 1988; Thomm & Wich, 1988). This region, known as the "distal promoter element" (DPE), has been studied

using various archaeal *in vitro* transcription systems. These systems require promoter-containing template DNA, purified RNA polmerase, and cell extract containing essential transcription factors (Hausner & Thomm, 1993; Hudepohl *et al.*, 1990; Frey *et al.*, 1990).

4.1.2 Halobacterial Consensus Promoters

There are two published consensus haloarchaeal promoters to date. One has the two conserved sequence elements, box A and box B, closely resembling the general consensus archaeal promoter (Hain *et al.*, 1992; Reiter *et al.*, 1988; Thomm & Wich, 1988; Zillig *et al.*, 1993), and the second is more specific to the haloarchaea and consists of an AT-rich region with similarity to the eukaryal TATA box. The two-sequence element consensus promoter (Palmer & Daniels, 1995), has a core



The preference for a pyrimidine residue in the 5′ position of the region and the exclusion of guanine and cytosine in the next four positions in the 3′ direction are defined characteristics shared by all efficient archaeal promoters. A 10-nucleotide purine rich sequence, located 5′ of the box A element, was also seen to be important for effective transcription. The second haloarchaeal promoter has a consensus core of

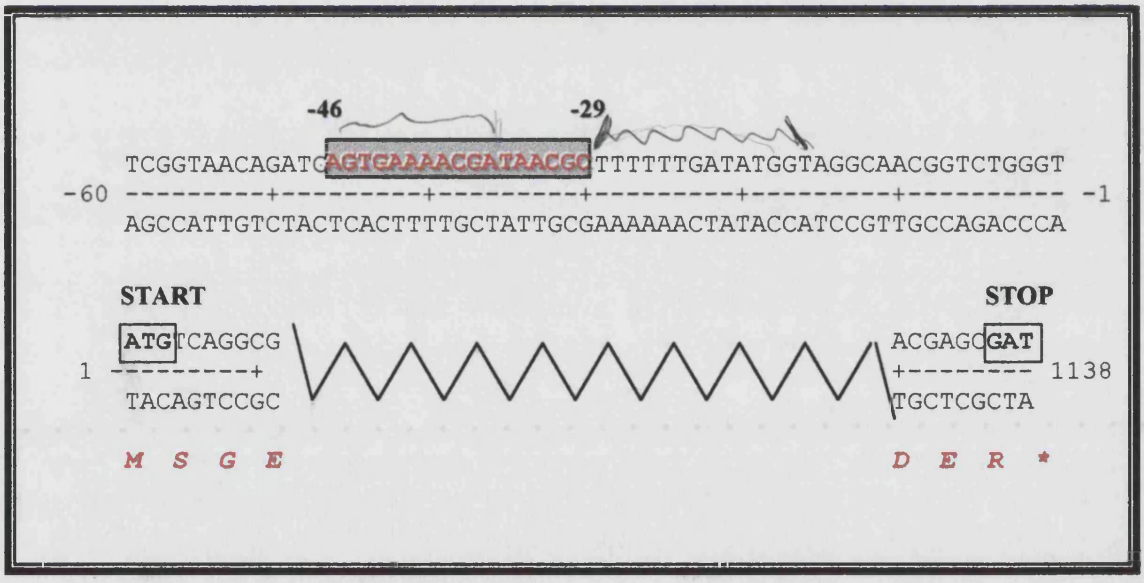
-32 (A/G)GT(A/T)(A/T)(A/T)(A/T)(A/G)AC (C/T)G(G/C)(C/T) –19

(Danner & Soppa, 1996).

4.1.3 Evidence of Regulatory Regions for the Citrate Synthase Gene

A search was made of the entire sequenced region for possible promoters and other regulatory regions. Since no other ORFs were detected in the sequenced 1 kb of non-citrate synthase coding DNA, it was assumed that the gene was not part of an operon system. This was no surprise as all other known citrate synthase genes are transcribed as individual units. A region closely resembling the AT-rich region of the consensus haloarchaeal promoter described by Danner & Soppa (1996) was detected 29 bp upstream of the start site (Fig 4.1).

Figure 4.1 Mapped Position of the Suspected Citrate Synthase Promoter



Numbers refer to nucleotide positions with respect to the gene sequence.
■ Indicates the position of the suspected haloarchaeal promoter.
Alignment was produced using MAPSORT (GCG).

In order to confirm that the above motif is the gene's promoter, promoter mapping experiments would have to be carried out. However, experiments described in chapter 6, involving the sequence thought to be the promoter, do give results consistent with this proposed role although the exact position can not be resolved from this evidence alone.

It is thought that in some organisms the genes involved in the citric acid cycle may cluster together within the genome. In the case of the citrate synthase gene, in excess of 500 nucleotides both up and downstream of the coding region have been sequenced and no ORFs have yet been detected indicating the absence of an operon arrangement.

The double stranded DNA sequenced included the position of the suspected promotor and the amino acid translation of the coding region of the citrate synthase gene can be seen in Figure 4.2.

4.2 Amino Acid Sequence Analysis

Citrate synthase from *Bacillus subtilis* and *Pyrococcus furiosus* were chosen as representatives with which to compare the amino acid composition of the *Haloferax* enzyme. *Bacillus subtilis* CS II is the closest matching homology score to Haloferax citrate synthase (see Table 4.2). *Pyrococcus furiosus* also displayed a high homology score to citrate synthase from *Haloferax* and it is also the closest matching archaeal enzyme for which there is a resolved crystal structure.

Figure 4.2 Complete Hf.volcanii Citrate Synthase Sequence

```
     GCGCCCGCCTTCACGACGGGACGGCAAAAGGCTTGCCAGGAGTAACGGTGCCACCGTGGT
-300 ---------+---------+---------+---------+---------+---------+ -241
     CGCGGGCGGAAGTGCTGCCCTGCCGTTTTCCGAACGGTCCTCATTGCCACGGTGGCACCA

     TCGAGAGAACTACGCCGTCGGGGGATATGAGTTTGCAGGTCGCCCGAGCGGCGCGCGCGC
-240 ---------+---------+---------+---------+---------+---------+ -181
     AGCTCTCTTGATGCGGCAGCCCCCTATACTCAAACGTCCAGCGGGCTCGCCGCGCGCGCG

     GACCCATCGGGGGGACTCCGTATAAGTGCCCCGAAAATAACGACAATCGTCGAACACTAT
-180 ---------+---------+---------+---------+---------+---------+ -121
     CTGGGTAGCCCCCCTGAGGCATATTCACGGGGCTTTTATTGCTGTTAGCAGCTTGTGATA

     TTTAGACACCGACTATCCGCGAAACCGACAGACGGGGCTGCCACTGTATGTGTATGGCGG
-120 ---------+---------+---------+---------+---------+---------+ -61
     AAATCTGTGGCTGATAGGCGCTTTGGCTGTCTGCCCCGACGGTGACATACACATACCGCC

                -46                 -29
     TCGGTAACAGATCAGTCAAAACGATAACCCTTTTTTGATATGGTAGGCAACGGTCTGGGT
 -60 ---------+---------+---------+---------+---------+---------+ -1
     AGCCATTGTCTACTCACTTTTGCTATTGCGAAAAAACTATACCATCCGTTGCCAGACCCA

     START
     ATGTCAGGCGAACTGAAGCGGGGGCTGGAAGGTGTGCTGGTCGCCGAATCGAAACTCAGT
   1 ---------+---------+---------+---------+---------+---------+ 60
     TACAGTCCGCTTGACTTCGCCCCCGACCTTCCACACGACCAGCGGCTTAGCTTTGAGTCA

     M  S  G  E  L  K  R  G  L  E  G  V  L  V  A  E  S  K  L  S   20

     TTCATCGACGGCGACGCGGGACAACTCGTTTACTGCGGGTACGATATCGAAGACCTCGCA
  61 ---------+---------+---------+---------+---------+---------+ 120
     AAGTAGCTGCCGCTGCGCCCTGTTGAGCAAATGACGCCCATGCTATAGCTTCTGGAGCGT

     F  I  D  G  D  A  G  Q  L  V  Y  C  G  Y  D  I  E  D  L  A   40

     cgGGACGCAAGTTACGAGGAAGTACTCTATCTCCTGTGGCACGGAGCGCTTCCGACGGGC
 121 ---------+---------+---------+---------+---------+---------+ 180
     gcCCTGCGTTCAatgCTCCTTCATGAGATAGAGGACACCGTGCCTCGCGAAGGCTGCCcG

     R  D  A  S  Y  E  E  V  L  Y  L  L  W  H  G  A  L  P  T  G   60

     GAGGAACTCGACGCGTTCTCCGACGAGCTCGCGGCCCACCGCGACCTCGACGACGGcGTC
 181 ---------+---------+---------+---------+---------+---------+ 240
     CTCCTTGAGCTGCGCAAGAGGCTGCTCGAGCGCCGGGTGGCGCTGGAGCTGCTGCCgCAG

     E  E  L  D  A  F  S  D  E  L  A  A  H  R  D  L  D  D  G  V   80

     CTCGACGTGGCACGCGAACTCGCCGAACAGGACGAGTCGCCGATGGCGGCGCTCCGAACG
 241 ---------+---------+---------+---------+---------+---------+ 300
     GAGCTGCACCGTGCGCTTGAGCGGCTTGTCCTGCTCAGCGGCTACCGCCGCGAGGCTTGC
```

```
    L  D  V  A  R  E  L  A  E  Q  D  E  S  P  M  A  A  L  R  T   100

    CTCGTCTCGGCGATGTCGGCGTACGACGAAAGCGCCGACTTCGAGGACGTGACCGACCGC
301 ---------+---------+---------+---------+---------+---------+ 360
    GAGCAGAGCCGCTACAGCCGCATGCTGCTTTCGCGGCTGAAGCTCCTGCACTGGCTGGCG

    L  V  S  A  M  S  A  Y  D  E  S  A  D  F  E  D  V  T  D  R   120

    GAGGTCAACCTCGAGAAGGCAAAGCGCATCACGGCGAAGATGCCGTCGGTGCTCGCGGCC
361 ---------+---------+---------+---------+---------+---------+ 420
    CTCCAGTTGGAGCTCTTCCGTTTCGCGTAGTGCCGCTTCTACGGCAGCCACGAGCGCCGG

    E  V  N  L  E  K  A  K  R  I  T  A  K  M  P  S  V  L  A  A   140

    TACGCCCGCTTCCGTCGCGGCGACGACTACGTCGAACCCGACGAGAGCCTGAATCACGCG
421 ---------+---------+---------+---------+---------+---------+ 480
    ATGCGGGCGAAGGCAGCGCCGCTGCTGATGCAGCTTGGGCTGCTCTCGGACTTAGTGCGC

    Y  A  R  F  R  R  G  D  D  Y  V  E  P  D  E  S  L  N  H  A   160

    gCGAACTTCCTCTACATGCTCAACGCCGAGGAGCCGAACGAGGTGTCTGCCGAGACGTTC
481 ---------+---------+---------+---------+---------+---------+ 540
    cGCTTGAAGGAGATGTACGAGTTGCGGCTCCTCGGCTTGCTCCACAGACGGCTCTGCAAG

    A  N  F  L  Y  M  L  N  A  E  E  P  N  E  V  S  A  E  T  F   180

    GACATGGCGCTCGTgCTCCACACCACCACGGACTTGAACGCCTCGACCTTCTCCGCGATG
541 ---------+---------+---------+---------+---------+---------+ 600
    CTGTACCGCGAGCAcGAGGTGTGGTGGTGCCTGAACTTGCGGAGCTGGAAGAGGCGCTAC

    D  M  A  L  V  L  H  T  T  T  D  L  N  A  S  T  F  S  A  M   200

    GTCACGTCCTCGACGCTCTCTGACCTCTACAGCGCAGTCACGTCCGCAATCGGCACGCTC
601 ---------+---------+---------+---------+---------+---------+ 660
    CAGTGCAGGAGCTGCGAGAGACTGGAGATGTCGCGTCAGTGCAGGCGTTAGCCGTGCGAG

    V  T  S  S  T  L  S  D  L  Y  S  A  V  T  S  A  I  G  T  L   220

    TCGGGGTCGCTCCACGGCGGCGCGAACGCGAACGTCATGCGGATGCTGAAGGACGTCGAC
661 ---------+---------+---------+---------+---------+---------+ 720
    AGCCCCAGCGAGGTGCCGCCGCGCTTGCGCTTGCAGTACGCCTACGACTTCCTGCAGCTG

    S  G  S  L  H  G  G  A  N  A  N  V  M  R  M  L  K  D  V  D   240

    GACAGCGACATGGACCCGACCGAGTGGGTCAAAGACGCCCTCGACCGCGGCGAGCGCGTC
721 ---------+---------+---------+---------+---------+---------+ 780
    CTGTCGCTGTACCTGGGCTGGCTCACCCAGTTTCTGCGGGAGCTGGCGCCGCTCGCGCAG

    D  S  D  M  D  P  T  E  W  V  K  D  A  L  D  R  G  E  R  V   260

    GCGGGGTTCGTCCACCGCGTCTACAACGTCAAGGACCCCCGCGCGAAGATTCTCGGCGCG
781 ---------+---------+---------+---------+---------+---------+ 840
    CGCCCCAAGCAGGTGGCGCAGATGTTGCAGTTCCTGGGGGCGCGCTTCTAAGAGCCGCGC

    A  G  F  V  H  R  V  Y  N  V  K  D  P  R  A  K  I  L  G  A   280

    AAGTCCGAGGCGCTCGGCGAGGCCGCCGGCGACATGAAGTGGTACGAGATGTCGGTCGCC
841 ---------+---------+---------+---------+---------+---------+ 900
    TTCAGGCTCCGCGAGCCGCTCCGGCGGCCGCTGTACTTCACCATGCTCTACAGCCAGCGG

    K  S  E  A  L  G  E  A  A  G  D  M  K  W  Y  E  M  S  V  A   300
```

```
     ATCGAGGAGTACATCGGCGAGGAGAAGGGCCTCGCGCCGAACGTGGACTTCTACTCCGCG
 901 ---------+---------+---------+---------+---------+---------+ 960
     TAGCTCCTCATGTAGCCGCTCCTCTTCCCGGAGCGCGGCTTGCACCTGAAGATGAGGCGC

     I  E  E  Y  I  G  E  E  K  G  L  A  P  N  V  D  F  Y  S  A   320

     TCGACGTACTACCAGATGGGCATCCCCATCGACcTCTaCacTCCCATCTTCGCCGTCTCG
 961 ---------+---------+---------+---------+---------+---------+ 1020
     AGCTGCATGATGGTCTACCCGTAGGGGTAGCTGgAGAtGtgAGGGTAGAAGCGGCAGAGC

     S  T  Y  Y  Q  M  G  I  P  I  D  L  Y  T  P  I  F  A  V  S   340

     CGCGCCGGCGGCTGGATTGCGCACGTCCTCGAACAGTACGAGGACAACCGCCTCATCCGC
1021 ---------+---------+---------+---------+---------+---------+ 1080
     GCGCGGCCGCCGACCTAACGCGTGCAGGAGCTTGTCATGCTCCTGTTGGCGGAGTAGGCG

     R  A  G  G  W  I  A  H  V  L  E  Q  Y  E  D  N  R  L  I  R   360
                                                          STOP
     CCCCGCGCCCGCTACACCGGCGAGAAAGACCTCGACTTCACGCCGGTCGACGAGCGAT
1081 ---------+---------+---------+---------+---------+--------- 1138
     GGGGCGCGGGCGATGTGGCCGCTCTTTCTGGAGCTGAAGTGCGGCCAGCTGCTCGCTA

     P  R  A  R  Y  T  G  E  K  D  L  D  F  T  P  V  D  E  R  *   380

     AGGCGACGAACCGCACGCGACTGTTTTCCGTCTGCTGCCCCCGGAGCGACGCGTCGTTCC
  +1 ---------+---------+---------+---------+---------+---------+ +60
     TCCGCTGCTTGGCGTGCGCTGACAAAAGGCAGACGACGGGGGCCTCGCTGCGCAGCAAGG


     CTTCACCCTCCGCCTCCCGCTCTCGCGTTCGCCGTCCGCCGCACGCGCCGCCCTCAGTCG
 +61 ---------+---------+---------+---------+---------+---------+ +120
     GAAGTGGGAGGCGGAGGGCGAGAGCGCAAGCGGCAGGCGGCGTGCGCGGCGGGAGTCAGC


     TCTTCGCCCACGTCGTCGACGCCGACGCTCCCGCGTTCGACCATGAGCGAGTCGGCGTGG
+121 ---------+---------+---------+---------+---------+---------+ +180
     AGAAGCGGGTGCAGCAGCTGCGGCTGCGAGGGCGCAAGCTGGTACTCGCTCAGCCGCACC


     CCCGCCCAGTCGGCGTAGGAGAACGCGACGAGCGCGAACAGCCCCATCCACGTGAACGAC
+181 ---------+---------+---------+---------+---------+---------+ +240
     GGGCGGGTCAGCCGCATCCTCTTGCGCTGCTCGCGCTTGTCGGGGTAGGTGCACTTGCTG


     AGGACGATGCCGATTCGCGCGCCGAGCGGGCCGTAGCCGAACCCCTCGACGAGCAGGTAC
+241 ---------+---------+---------+---------+---------+---------+ +300
     TCCTGCTACGGCTAAGCGCGCGGCTCGCCCGGCATCGGCTTGGGGAGCTGCTCGTCCATG
```

Numbers refer to nucleotide positions with respect to the gene sequence.
Amino acid sequnce and numbering is indicated in red.
Start and stop codons are shown in bold.
The suspected promoter is shown as white font on a grey background.

Table 4.1 Comparison of Amino Acid Breakdown Compared to that of Bacillus subtilis and Pyrococcus furiosus

Amino Acid Code	Amino Acid Name	No.of AA			% AA		
		Hf.vol	Py.fu	Ba.su	Hf.vol	Py.fu	Ba.su
BASIC							
K - Lys	Lysine	13	33	22	3.4	8.8	6.0
R - Arg	Arginine	21	17	20	5.5	4.5	5.5
H - His	Histidine	8	8	14	2.1	2.1	3.8
ACIDIC							
D - Asp	Aspartic Acid	37	11	14	9.8	2.9	3.8
E - Glu	Glutamic Acid	35	39	38	9.2	10.4	10.4
UNCHARGED POLAR							
N - Asn	Asparagine	11	11	11	2.9	2.9	3.0
Q - Gln	Glutamine	4	5	8	1.1	1.3	2.2
S - Ser	Serine	26	19	21	6.9	5.1	5.8
T - Thr	Threonine	15	15	23	3.6	4.0	6.3
Y - Tyr	Tyrosine	18	25	16	4.7	6.6	4.4
NONPOLAR							
G - Gly	Glysine	27	29	23	7.1	7.7	6.3
A - Ala	Alanine	45	28	40	11.8	7.4	11
V - Val	Valine	27	22	20	7.1	5.9	5.5
L - Leu	Leucine	39	37	34	10.3	9.8	9.3
I - Ileu	Isoleucine	12	33	21	3.2	8.8	5.8
P - Pro	Proline	12	17	14	3.2	4.5	3.8
F - Phe	Phenylalanine	11	10	11	2.9	2.7	3.0
M - Met	Methionine	13	9	11	3.4	2.4	3.0
W - Trp	Tryptophan	4	7	2	1.1	1.9	0.5
C - Cys	Cysteine	1	1	2	0.3	0.3	0.5

Comparison made using PEPTIDESORT (GCG).

It has long been recognised that the bulk protein from the extreme halophiles is acidic in nature, the content of aspartate and glutamate residues being as high as 20% (Lanyi, 1974; Reistad, 1970). This phenomenon is once again observed, with the *Haloferax* CS where these acidic residues form 19% of the total gene sequence. This percentage is high when compared to citrate synthase from its non-halophilic counterparts, which have an acidic amino acid content of 14% and 13%, respectively. It should be noted, however, that in all three cases the number of glutamate residues remains similar, with the non-halophilic proteins containing slightly more, and that the increase in acidity is due to the greater number of aspartate residues (3-fold) present in *Haloferax* CS.

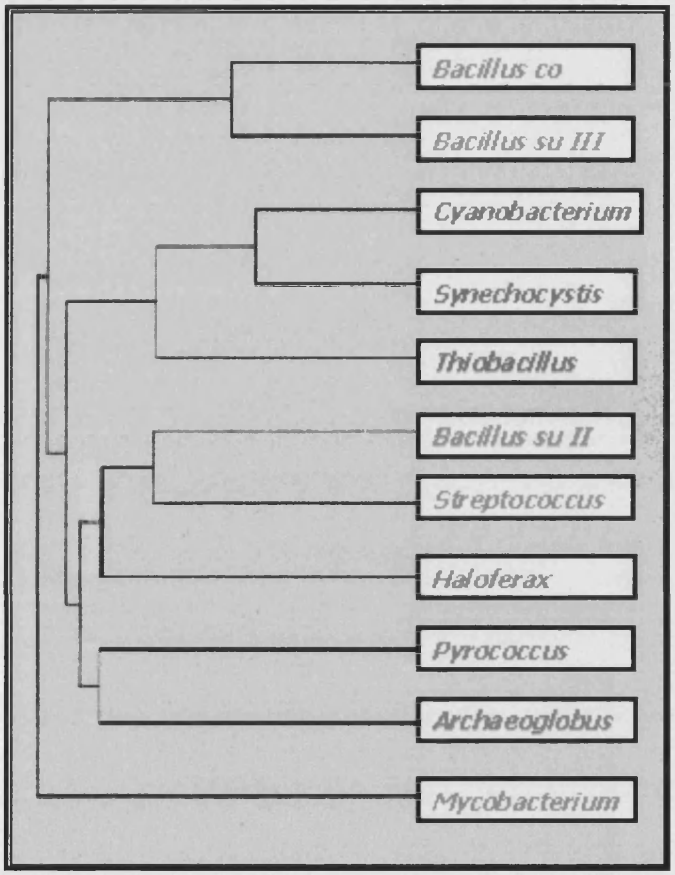
4.2.1 Homology Scores

A comparison of the closest matching sequences in the EBI protein database was achieved by performaing a FASTA3 homology search. The closest homology match is to citrate synthase II from the Gram positive bacterium, *Bacillus subtilis*, with 50% amino acid identity over a 373 overlap; the next closest was citrate synthase from the Archaeon *Archaeoglobus fulgidus* with a 48% sequence identity over a 378 amino acid overlap. The ten best matched homology scores to *Haloferax volcanii* citrate synthase are shown in Table 4.2, these homology scores are put into a pictorial context in Figure 4.3 in the form of a dendrogram.

Table 4.2 Top Ten Homology Scores for Hf. volcanii Citrate Synthase Gene

Organism	% identity (amino acid overlap)
Bacillus subtilis II	50% (373)
Archaeoglobus fulgidus	48% (378)
Cyanobacterium anabaena	44% (386)
Streptococcus mutans	44% (374)
Synechocystis	43% (379)
Pyrococcus furiosus	42% (378)
Thiobacillus ferroxidans	41% (367)
Bacillus coagulans	40% (367)
Bacillus subtilis III	40% (365)
Mycobacterium smegmatis	39% (371)

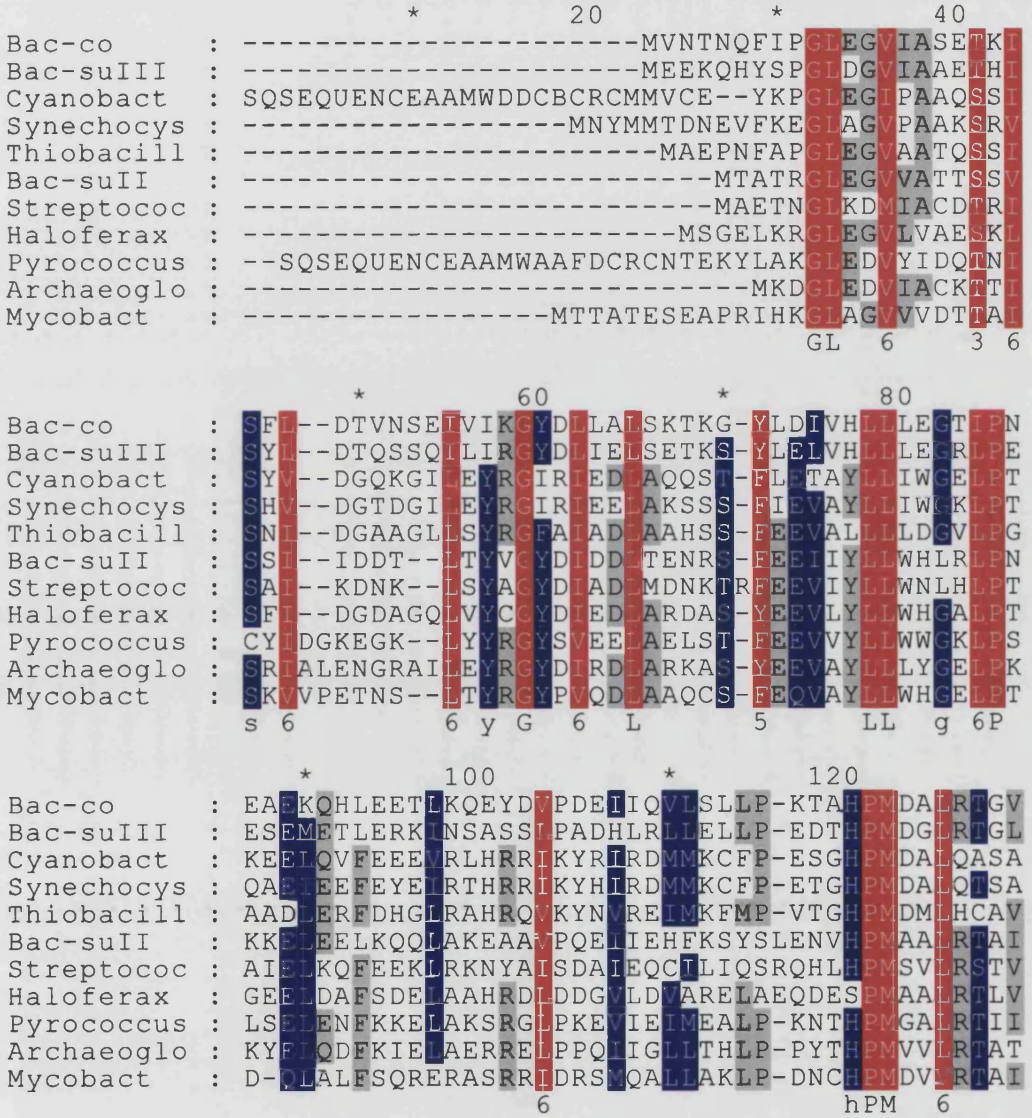
Figure 4.3 Dendrogram Showing the Relatedness of the Ten Best Matched Protein Sequences to Hf. volcanii Citrate Synthase

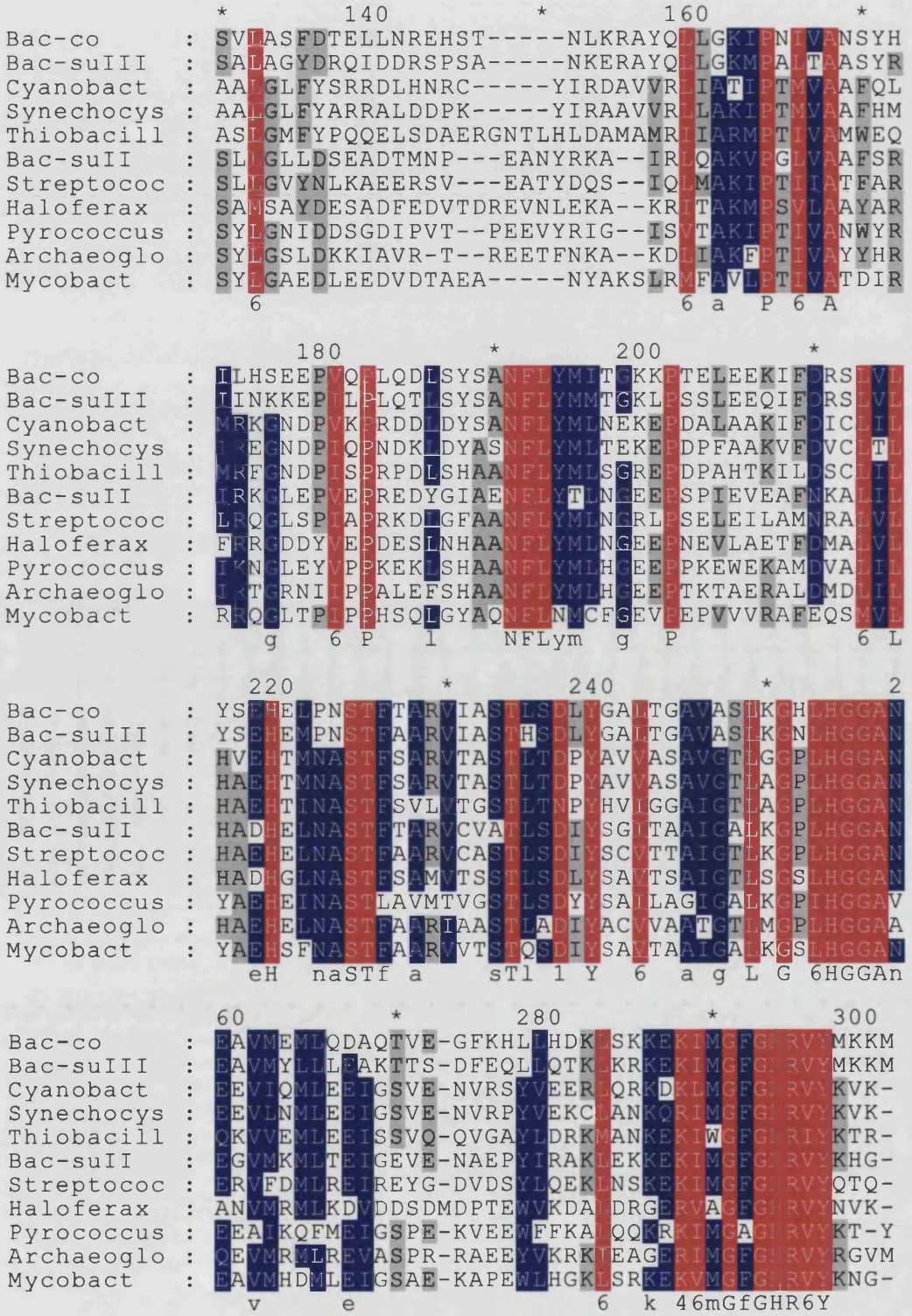


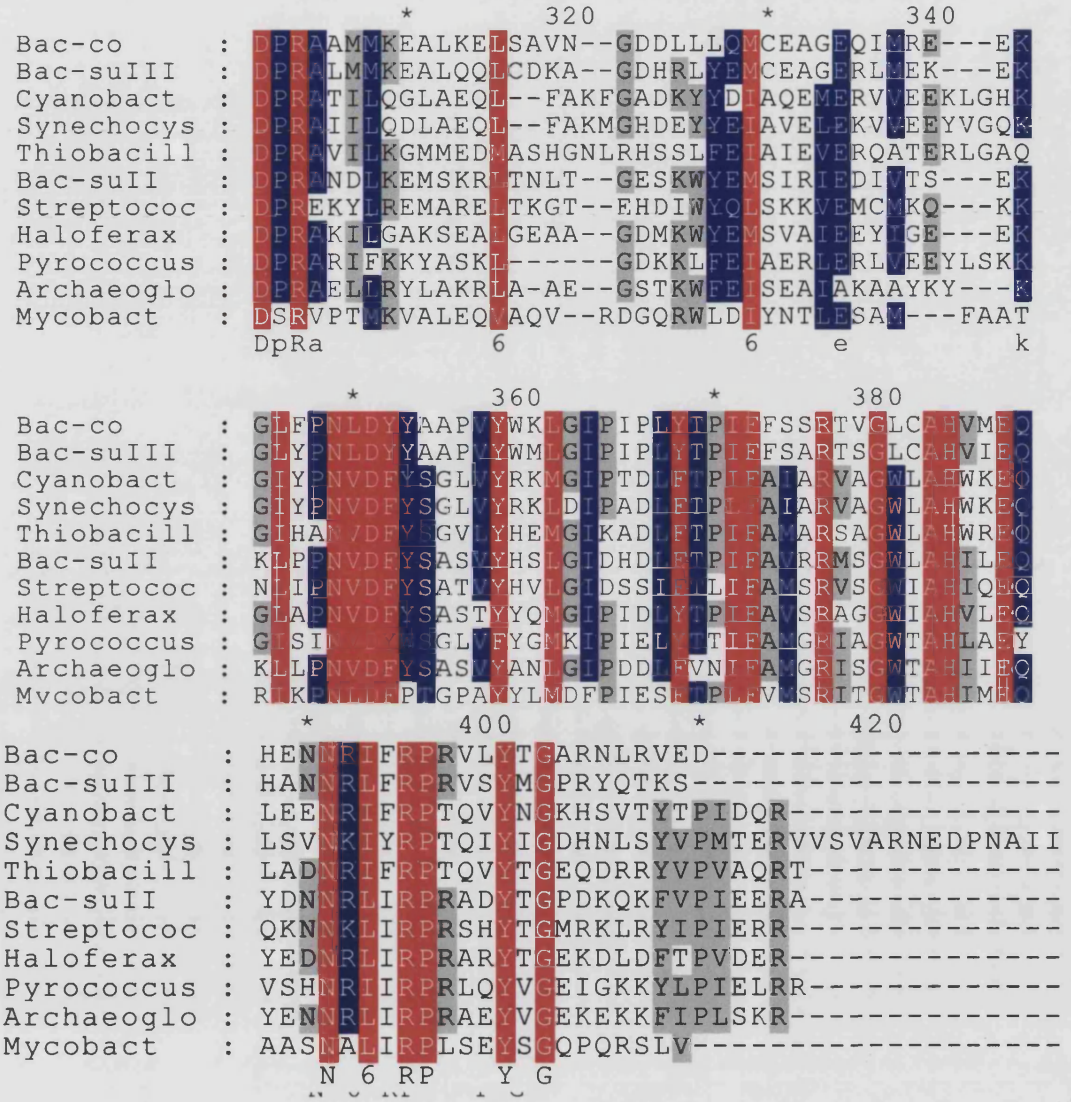
Produced using PILEUP and FIGURE (GCG)

The high level of sequence identity between the CS genes of the organisms shown in Table 4.2 is put in to context in Figure 4.4. The protein sequences from the organisms shown are also used for a further alignment with additional citrate synthase sequences. This additional alignment has subsequently been used for the construction of phylogenetic trees and will be discussed later in this chapter. A table containing the full names of the organisms used in this and the subsequent 'pileup' can be found in appendix 2.

Figure 4.4 Alignment of the ten best matched sequences to Haloferax citrate synthase




Bac-co
Bac-suIII
Cyanobact
Synechocys
Thiobacill
Bac-suII
Streptococ
Haloferax
Pyrococcus
Archaeoglo
Mycobact



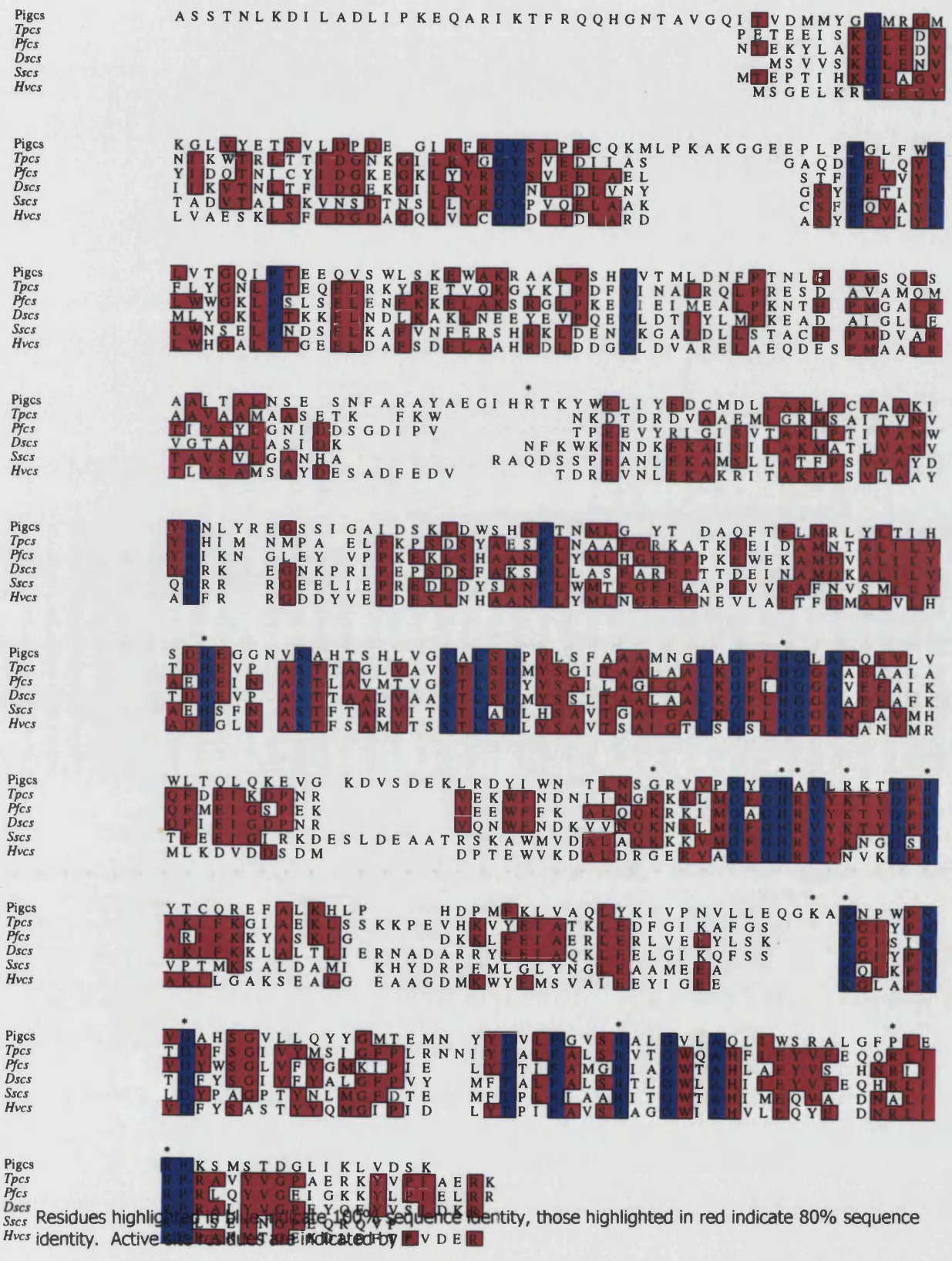
Produced using PILEUP (GCG). Presentation and shading of conserved regions produced using GENEDOC. Residues highlighted in red indicate 100% conservation in this particular line-up, blue indicates 80% identity and grey 60% identity. The full names of the organisms used in this alignment can be found in appendix 2.

4.3 Structural comparison

A structural alignment of the *Haloferax* citrate synthase has been conducted with citrate synthase protein sequences from Pig (37°C), *Thermoplasma* (55°C), *Pyrococcus* (100°C), DS23R (15°C) and *Sulfolobus* (85°C) using the programmes PILEUP (GCG) and MOLSCRIPT (Fig 4.5).

The active site residues denoted by * are highly conserved across the field of organisms shown. Of the 14 active site residues, 8 are 100% conserved across the field, with only 2 of the *Haloferax* residues differing from those of *Pyrococcus*. Despite having a high proportion of negatively charged amino acids and an overall negative electrostatic charge the *Haloferax* CS active site has an intense cluster of positively charged amino acid residues, this is due to the overall negative charge of the enzyme's substrates acetyl CoA and oxaloacetate. A picture of the citrate synthase active site of the organisms shown in this alignment can be seen in chapter 5 (Fig 5.7).

Figure 4.5 Structural Alignment of Citrate Synthases



Residues highlighted in blue indicate 100% sequence identity, those highlighted in red indicate 80% sequence identity. Active site residues are indicated by *

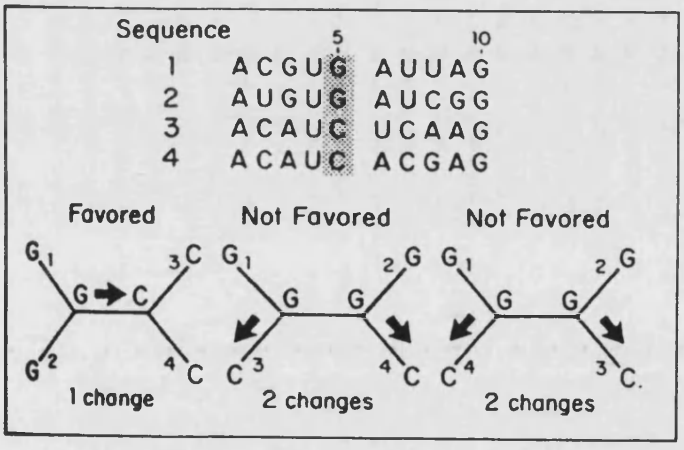
4.4 Phylogenetic Analysis

Phylogenetic analysis was carried out on the *Haloferax* citrate synthase and other known citrate synthase sequences using Felsenstein's Phylip package. Two different algorithms were chosen, Maximum Parsimony and Distance Matrix, both of which are commonly used in this type of analysis and are included in the Phylip package.

4.4.1 Maximum Parsimony Algorithm

Parsimony is a character-based analysis, that uses the information provided by each nucleotide and selects the tree that is derived from the minimum number of nucleotide changes. An example using DNA analysis, clearly illustrates the method.

Figure 4.6 An Illustration of Parsimony



The pattern GGCC, from position 5 of the four sequences could be obtained by three different arrangements of the organisms 1,2,3 and 4, as demonstrated by the three tree diagrams in figure 4.6. Parsimony would support the placement of organism 1 with organism2 and organism 3 with 4, as this arrangement allows the observed pattern, GGCC, to be obtained via 1 change whereas the other arrangements of the organisms each require 2 changes. The most parsimonious tree supports the minimum number of

changes. For a set of sequences 500 nucleotides long, the method performs this analysis at each position and selects the tree supported by parsimony at the majority of positions (Lake, 1989).

4.4.1.1 Protpars

Protpars is a parsimony algorithm for protein sequences which counts the minimum number of nucleotide changes required to produce the observed amino acid changes within the given protein sequences, and supports the tree with the minimum requirement of change. This method only allows changes of amino acids that are supported by the genetic code; however some amino acid changes that should not occur according to this rule and would therefore be incorrect are nevertheless allowed, by permitting a change between two amino acids via a third, providing each change is supported by the genetic code. For example, it is possible to change phenylalanine to glutamine via leucine:- UUU (phe) -> CUU (Leu) -> CUA (Leu) -> CAA (Glu). This substitution would only count as two steps as this programme does not count synonymous changes, which have been shown to be considerably faster and easier than ones that change the amino acid. For this reason the method assumes that synonymous changes do not need to be counted, thus the change between the two leucine codons is ignored.

The main assumptions of this method are

1) Change in different sites is independent
2) Change in different lineages is independent
3) The probability of a base substitution that changes the amino acid sequence is small over the lengths of time involved in a branch phylogeny.
4) The expected amounts of change in different branches of the phylogeny do not vary by so much that two changes in a high-rate branch are more probable than one change in a low rate branch.

5) The expected amounts of change do not vary enough among sites that two changes in one site are more probable than one change in another.
6) The probability of a base change that is synonymous is much higher than the probability of a change that is not synonymous.

Parsimony is not able to detect homoplassy (convergent or parallel evolution) and is affected by unequal rates of evolution, which causes long branch organisms (fast evolving) to be grouped together and short branch organisms (slow evolving) to be grouped together, predicting an incorrect topology.

4.4.2 The Distance Matrix Algorithm

Two programmes are required to produce a distance matrix tree; one to construct the distance matrix (Protdist) and the other to produce a tree from the matrix (Fitch).

4.4.2.1 Protdist

Protdist computes a distance matrix of protein sequences by performing pairwise comparisons and calculating the number of amino acid substitutions between these two sequences. It is assumed that the greater the number of amino acid substitutions present at the time of comparison, the greater the evolutionary divergence is between the two sequences. Pairwise comparisons continue until all possible pairs of sequences have been compared, producing a table likeness to Table 4.3 below. Table 4.3 illustrates the values that can be calculated from the pairwise comparisons, showing that the organisms A and B, with a value of 0.08 are more closely related than organisms A and C, with a value of 0.19 and hence there are more amino acid substitutions between A and C than there are between A and B.

The calculated distances are scaled in terms of the expected number of amino acid substituions per site but each site is assumed equal.

Table 4.3 A Hypothetical Distance Matrix Table

	A	B	C	D	E
A	-	0.08	0.19	0.70	0.65
B		-	0.17	0.75	0.70
C			-	0.80	0.60
D				-	0.12
E					-

This method only utilises the differences it can directly observe between the sequences and is therefore not able to account for certain underlying evolutionary events. Firstly, this method is unable to detect convergent or parallel evolution where identical sites of amino acids between two sequences are achieved by chance. Secondly, the different rates of evolution at different sites within the protein, e.g. conserved and unconserved regions , are overlooked. Thirdly, individual nucleotide substitutions are often not detected due to the degeneracy of the DNA code, where a mutation in the third nucleotide of the codon on occasions does not alter the amino acid coded for, e.g. phenylalanine can be coded for by UUU or UUC. Fourthly, multiple replacement of an amino acid at the same site during evolution is unaccounted for.

The inability of the method to detect these underlying events means that information about the behaviour of individual sequence positions is fortified and often causes the method to underestimate the number of evolutionary changes between two sequences. Despite these faults this method is upheld as surprisingly effective when dealing with real data and hence is widely used.

4.4.2.2 Fitch

Fitch uses the distance matrix computed by Protdist to produce a phylogenetic tree. It is based on the Fitch-Margoliash method and estimates branch lengths between the organisms based on the values from the distance matix table, to create the most likely phylogenetic arrangement of all of the organisms. The approach is most easily understood by this illustration, using the values from Table 4.3 to produce the branches in Figure 4.7.

Figure 4.7 Illustration of the Branch Lengths of Organisms A,B and C

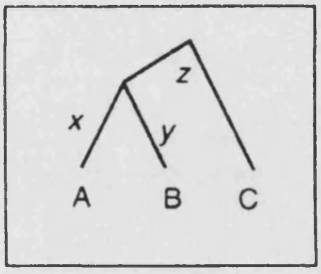


Figure 4.7 shows the branch lengths x,y and z which separate the organisms A,B and C. From Table 4.3 it is known that, x + y = 0.08; x + z = 0.19; and y + z = 0.17. By definition distance AB = x + y, distance AC = x + z, distance BC = y + z ; therefore, the branch lengths can be calculated by solving simultaneous equations:

(x + y = 0.08)

(x + z = 0.19)

2y + y + z = 0.27)

(y + z = 0.17)

2x = 0.10 or **x = 0.05**

Therefore x = 0.05, y = 0.03, and z = 0.14, which are branch lengths. This process is continued until the tree is complete. In this example the branch lengths are additive but this would not be the case with a larger number of organisms.

4.4.3 Construction of Phylogenetic Trees

The trees were constructed from a multiple alignment performed using PILEUP (GCG). The multiple alignment on which the trees are based can be seen in appendix 3. The multiple alignment was transformed into a format that phylip can read using CLUSTALW(1.5). The output file from CLUSTALW was read into SEQBOOT to select the parameters for bootstrapping. The output file from this programme was in turn read into the appropriate algorithm files PROTPARS or PROTDIST. The output file from PROTDIST had to be run in FITCH in order for the algorithm to be converted to a tree. The files from either protpars or fitch, depending on the algorithm selected, were run in CONSENSE to combine the information obtained into a single consensus tree. The output from this programme in the form of TREEFILE was converted into a tree image by the use of DRAWTREE.

All citrate synthase sequences available on the SWISSPROT database, or sequences showing a high degree of similarity to known citrate synthase sequences from published complete genomes, were selected for analysis. Both trees produced were constructed from the same multiple alignment. Figure 4.8 is a pictorial representation illustrating the relatedness of the citrate synthase proteins from the organisms chosen. A table of the full names of the organisms selected can be seen in appendix 3, and the multiple sequence alignment on which the phylogenetic trees were constructed in appendix 3.

Figure 4.8 Dendrogram showing the Relatedness of the Citrate Synthases Selected for Phylogenetic Analysis

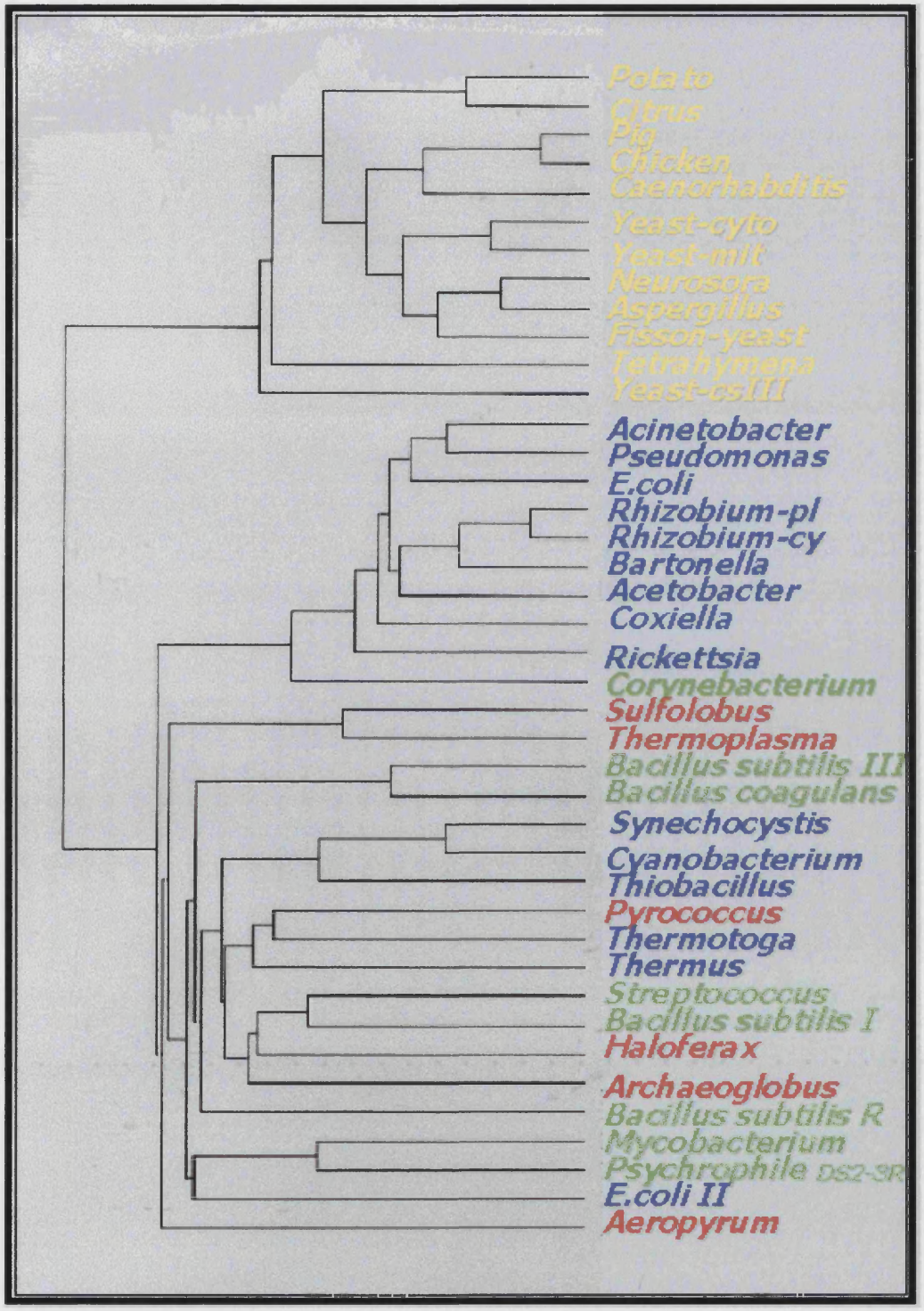


Figure produced using PILEUP and FIGURE (GCG)

Eukaryotes **Gram Positive Bacteria** **Gram Negative Bacteria** **Archaea**

4.4.4 Distance Matrix Phylogenetic Tree

The distance matrix tree (Fig 4.9) shows three clearly defined groups, one consisting of Eukaryotes, another of the Archaea and Gram-positive bacteria and a third of Gram-negative bacteria. There are a few exceptions to the general domain layout: *Corynebacterium* has grouped at the base of the Gram-negative branch, and *Thermus* and *Thermotoga* have grouped with *Pyrococcus* in the Archaeal/Gram-positive branch along with *E.coli II*. *Bacillus subtilis II* and *Bacillus R* have grouped between the Gram-negative bacteria and the Gram-positive/archaeal branch. Phylogenetic trees have been produced a number of times for this enzyme, each time being updated with the latest sequences. Each time they have been produced, *Corynebacterium* always appears within the Gram-negative branch; currently we have no obvious explanation for this.

Although taxonomically *Thermus* and *Thermotoga* have been defined as Gram-negative bacteria, it may not be too surprising that their citrate synthases cluster with that from their fellow hyperthermophile *Pyrococcus*. *Haloferax* groups within the archaeal cluster which tends to be placed at the base of the archaeal/Gram-positive branch, indicating that perhaps they are the ancestral predecessors. In order to confirm this assumption the tree needs to be rooted.

E.coli II citrate synthase has been defined as a 2-methyl citrate synthase by its ability to utilise propionyl-CoA as a substrate (Gerike *et al.*, 1998). The CS from the psychrophilic isolate DS23R has also been shown to have activity with propionyl-CoA (Gerike *et al.*, 1997) indicating that it too has 2-methyl citrate synthase activity. Perhaps the grouping of *E.coli II* with the psychrophilic isolate within a monophyletic branch could indicate the presence of a second type of citrate synthase, 2-methyl citrate synthase. The citrate synthases from other organisms clustering in this group, *Mycobacterium* and *Aeropyrum,* have never been tested for their ability to react with propionyl-CoA.

Figure 4.9 Consensus Distance Matrix Phylogenetic Tree

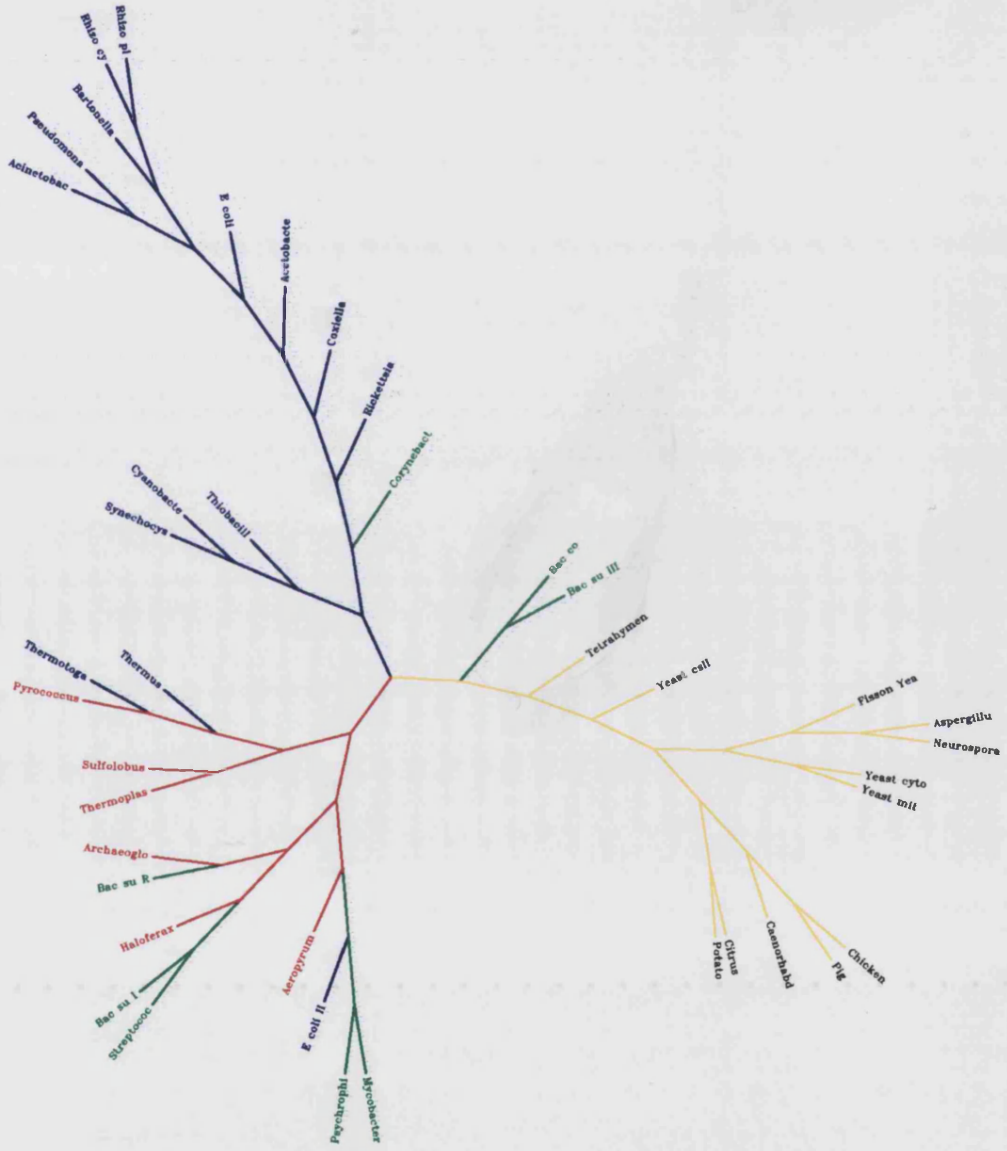


Figure produced using *PROTDIST, FITCH* and *DRAWTREE*

Eukaryotes **Gram Positive Bacteria** **Gram Negative Bacteria** **Archaea**

4.4.5 Parsimony Phylogenetic Tree

As with the distance matrix tree, the three main branches are clearly seen, in the parsimony tree (Fig 4.10), although they are not quite as well defined. Once again *Corynebacterium* has grouped at the base of the Gram-negative branch, *Thermotoga* and *Thermus* have grouped with *Pyrococcus* and *E.coli II* has clustered with the psychrophile isolate within the archaeal/Gram-positive branch. However, *Cyanobacterium, Synechocystis* and *Thiobacillus* have also grouped within the Gram-positive/archaeal branch. As before, *Haloferax* has grouped within the Archaea, which with the exception of *Aeropyrum* is condensed around the base of the Gram-positive/archaeal branch. The grouping of *E.coli II* once again with the pschrophilie isolate , *Mycobacterium* and *Aeropyrum*, gives credibility to the hypothesis that they are indeed 2-methyl citrate synthases, and this would also explain the positioning of the archaeon *Aeropyrum* at the end rather than at the base of the branch.

The consistent branching of *Pyrococcus* with the Gram-negative bacteria *Thermus* and *Thermotoga* in both trees, forming a monophyletic group close to the base of the archaeal/Gram-positive branch, could possibly be due to a horizontal gene transfer event between an Archaeon and a Gram-negative bacterium, or even an early evolutionary branch, although the latter explanation can not be given any credibility without the evidence of a rooted tree.

Figure 4.10 Consensus Parsimony Phylogenetic Tree

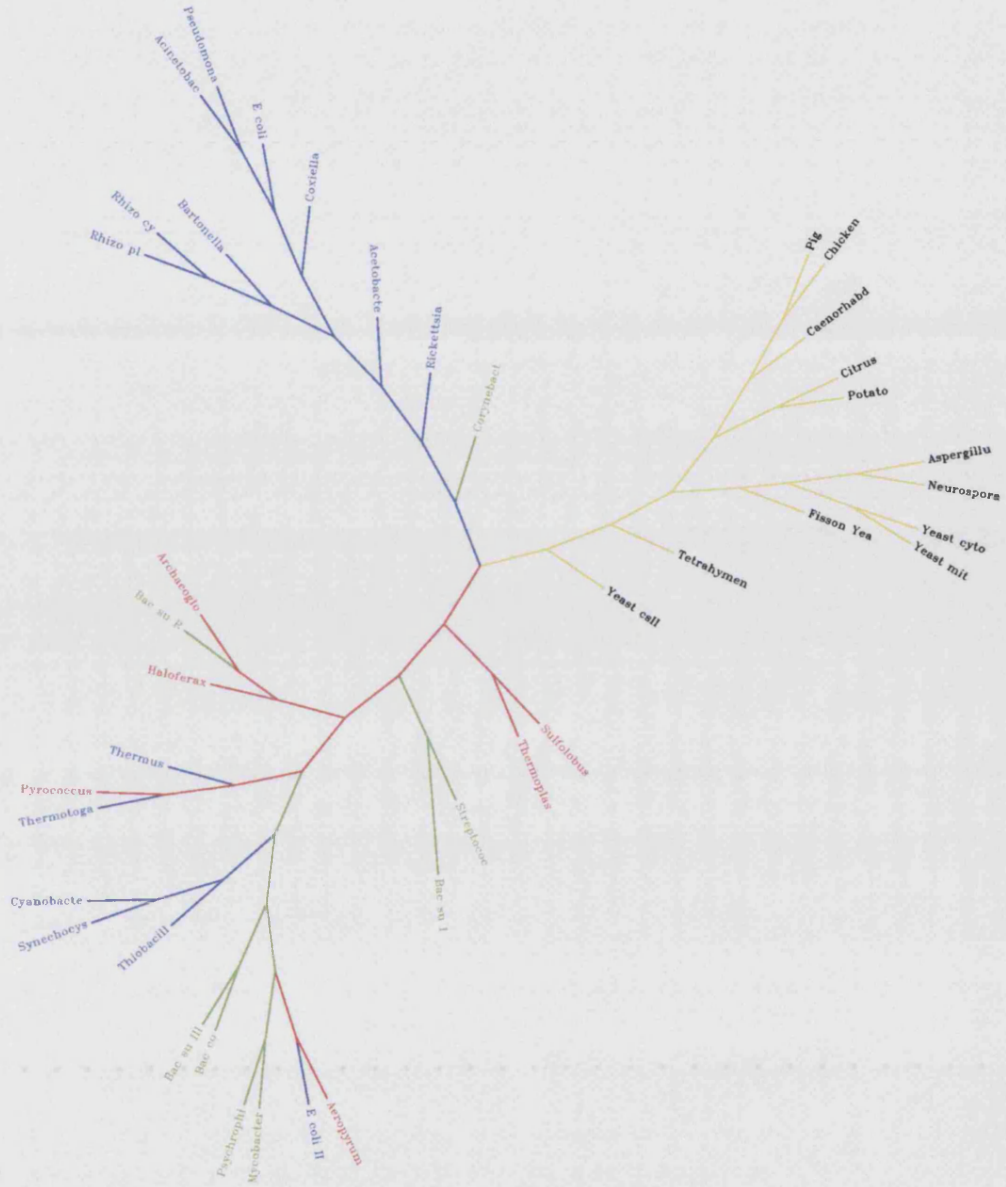


Figure produced using *PROTPARS* and *DRAWTREE*

Eukaryotes **Gram Positive Bacteria** **Gram Negative Bacteria** **Archaea**

4.5 Conclusions

The gene sequence itself has not revealed anything unusual. The gene initially appears to transcribe off its own promoter rather than be part of an operon system. The gene starts amino acid translation from a classic methionine ATG start codon, this may seem an obvious fact but other *Haloferax* genes have been known to start protein translation from valine (CCG) (Jolley *et al.*, unpublished).

The *Haloferax* citrate synthase amino acid sequence, when compared to other sequences in the database, has the highest sequence identity scores to citrate synthases from Gram-positive Bacteria and other Archaea. This is concurrent with initial analyses performed on the gene fragment (Seedhouse, 1995), and comparisons of other halophilic proteins (Jolley, 1996).

The protein sequence has an increased number of negatively-charged amino acid residues when compared to its non-halophilic counterparts. This increase in acidic residues is a classic feature of other halophilic gene sequences and structures. From our knowledge of the four halophilic protein structures currently available, the increased number of acidic amino acid residues will probably be put into a more meaningful context in a structural molecular model. This will be discussed further in chapter 5.

Phylogenetically the *Haloferax* citrate synthase is placed within the Archaea/Gram-positive branch which is where it was expected to be placed based on 16S rRNA trees and preliminary trees constructed using the partial citrate synthase gene fragment. Both trees constructed have placed *Haloferax* fairly close to the central branching inferring that it could be one of the earlier forms of the Archaea. No firm conclusions regarding ancestral lineages can be drawn without the presence of a rooted form of the tree, and this is difficult to produce without an obvious citrate synthase outlyer.

Chapter 5

Construction of a Molecular Model of Haloferax volcanii Citrate Synthase Based Upon the Crystal Structure of the Enzyme from Pyrococcus furiosus

5.1 Homology Model

Using the crystal structure of the citrate synthase from *Pyrococcus furiosus,* we have constructed an homology modelled structure of the halophilic enzyme. *Pyrococcus furiosus* was chosen as closest archaeal relative to *Hf volcanii* for which we have a resolved citrate synthase crystal structure (Russell *et al.*, 1997). *Pyrococcus* CS has 45% sequence identity and 65% similarity at the amino acid level to *Haloferax* CS making it an ideal candidate on which to base a model structure.

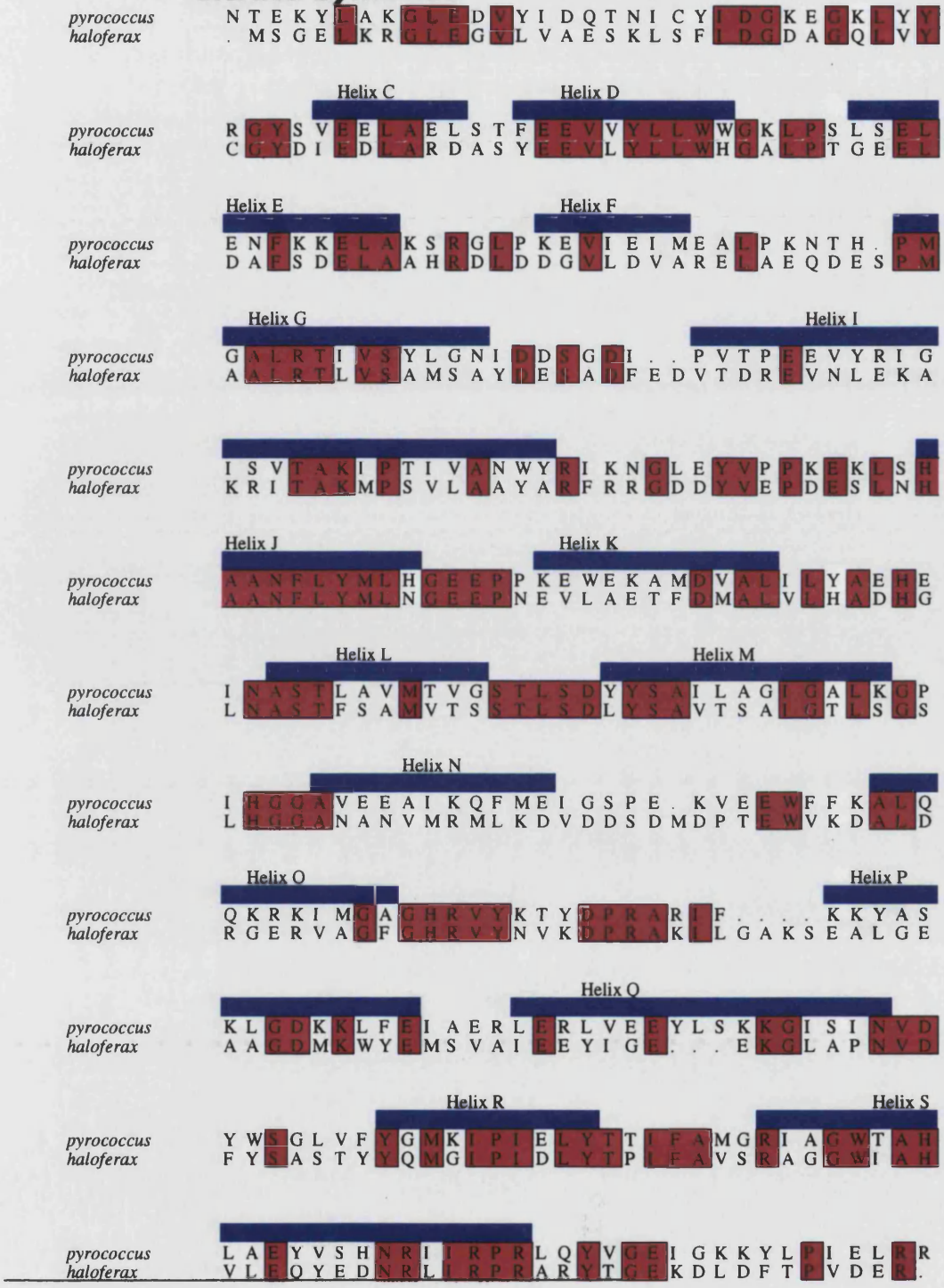
The model was constructed by performing a BESTFIT comparison using the GCG package to align the two gene sequences (Fig 5.1). Amino acid side chain residues were then substituted in the respective positions in the *Pyrococcus* structure to recreate a model of the *Haloferax* enzyme. The *Haloferax* protein sequence is 3 amino acid residues longer than its *Pyrococcus* counterpart; therefore, certain amino acids were shifted with respect to helix position so that they were introduced into a loop region of the structure rather than in the middle of a helix.

5.2 Three-Dimensional Structure

Figure 5.2 shows a ribbon structure of the modelled *Hf volcanii* citrate synthase. The two domains of each monomer are clearly visible in contrasting colours. The dimer interface towards the centre of Fig 5.2 is also highlighted with the use of separate colours to denote the large domain of each monomer. The active site residues are highlighted in a ball and stick representation.

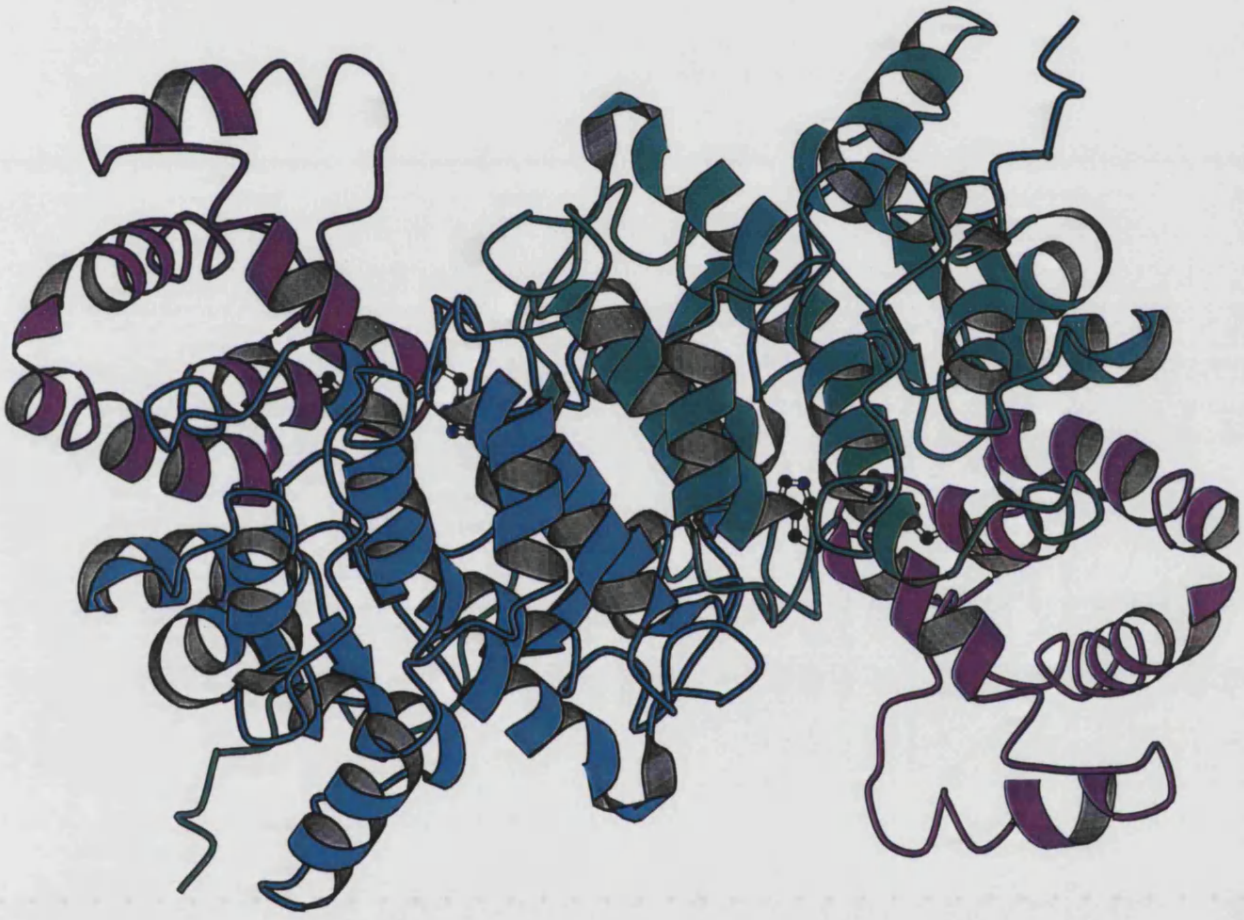
The ribbon structure of the modelled enzyme is useful for visualising the model as a whole, with respect to the positions of the two domains within the subunits, and for studying the dimer interface interactions. However, for the purpose of further investigation into the structural basis of halophilicity, a surface electrostatic profile of the enzyme is required.

Figure 5.1 Structural Alignment of Haloferax and Pyrococcus Citrate Synthase



Identical amino acids are highlighted in red
The position of the helices are shown in blue

Figure 5.2 Ribbon Structure of the H.volcanii citrate synthase model



The image was produced using the software programme 'O'.
The large domain of each monomer is shown in green and turquoise to highlight the dimer interface. Purple indicates the small domain of each monomer and the active site residues are depicted in the ball and stick representation.

5.3 Surface Profile

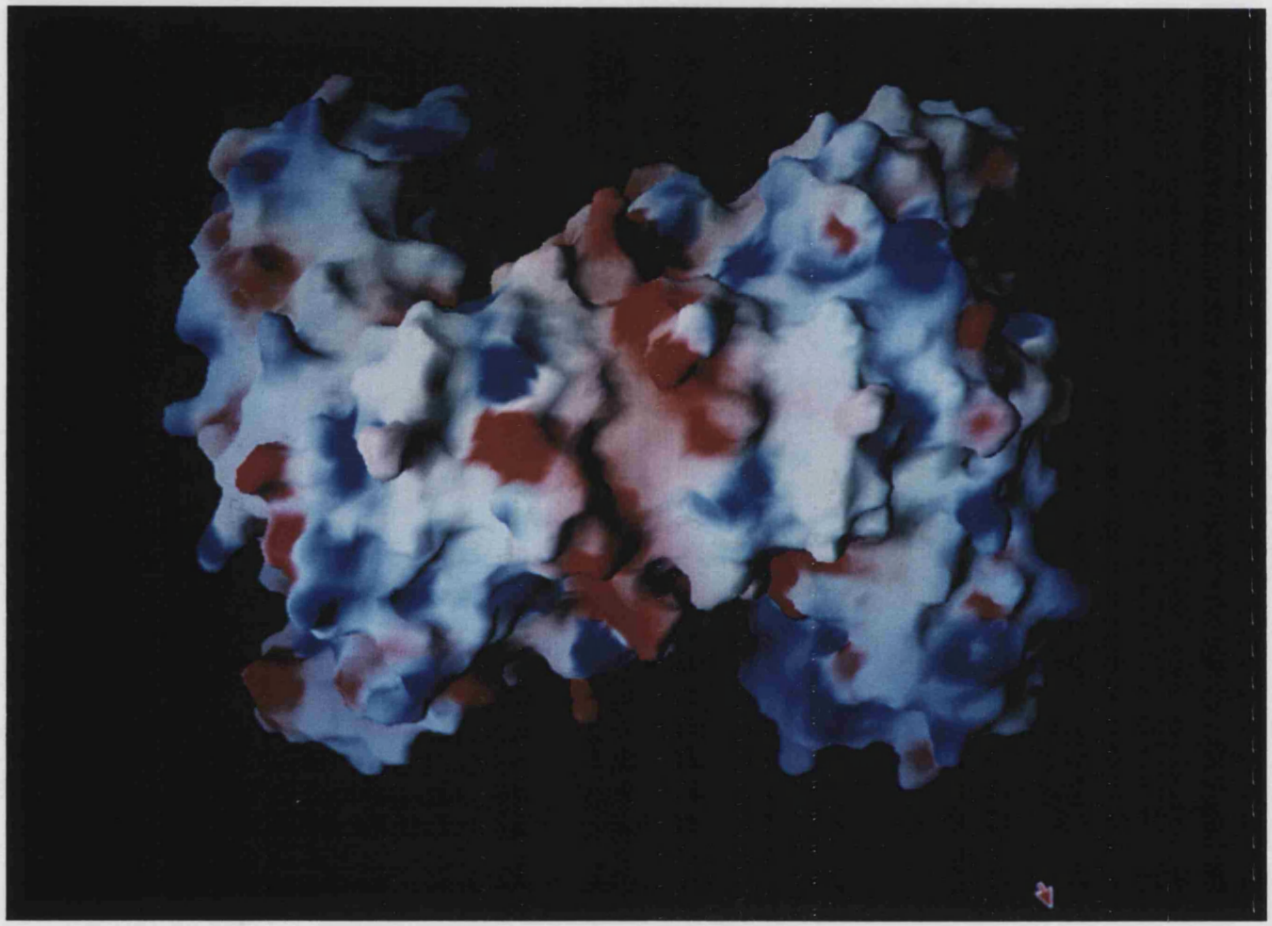
A molecular surface profile of the modelled citrate synthase has been produced emulating various physiological conditions. The electrostatic surface potential was calculated and displayed using the programme GRASP (Nicholls *et al.*, 1993). For comparison, the modelled enzyme surface profiles have been displayed in conjunction with those of *Pyrococcus* citrate synthase, the non-halophilic enzyme on which the molecular model of the halophilic protein was based.

The surface profile of the *Pyrococcus* citrate synthase in 0 M KCl, shows the predicted profile of a typical hydrated protein with a mixture of negative and positive charges interspersed with neutral regions (Fig 5.3).

When we simulate the same low salt conditions for the *Haloferax* modelled enzyme (Fig 5.4), it is clear that the surface potential is drastically different from that of *Pyrococcus*. The excess of negatively charged amino acids present in the *Haloferax* enzyme is clearly mostly on the surface of the protein, where they are in number at least twice those of the positive amino acids. Thus the net surface charge of the modelled enzyme is negative (-53), compared with the more neutral surface potential seen with the *Pyrococcus* enzyme (+15) (Table 5.2).

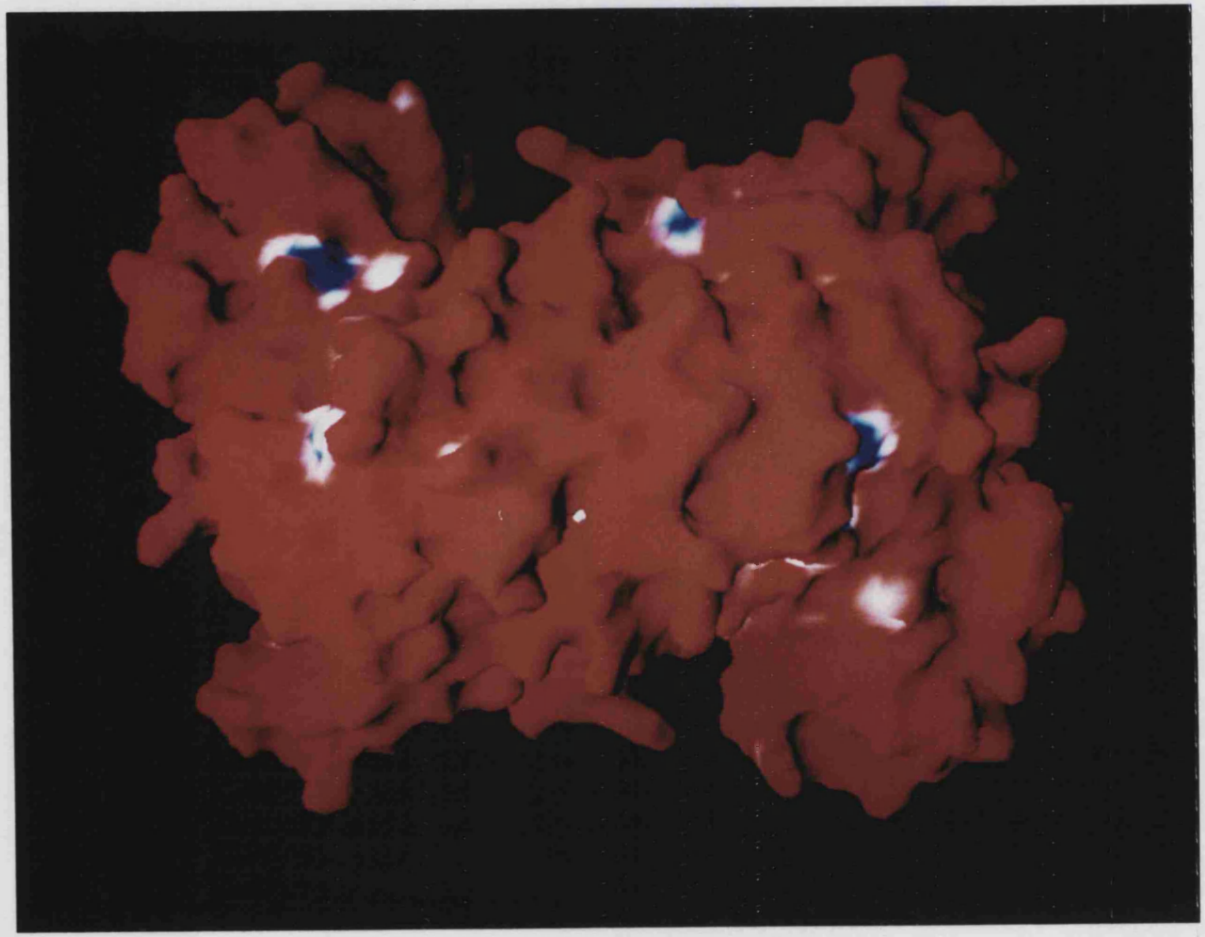
When conditions of 0.6 M KCl are simulated (Fig 5.5), there is a more balanced overall distribution of positive and negative potentials on the surface of the *Haloferax* CS due to the screening of the excess negative charges by hydrated K^+ ions. The surface profile now resembles that of the *Pyrococcus* citrate synthase in the absence of salt (KCl) (Fig 5.3). Increasing the concentration of salt to 2M KCl does not significantly change the surface electrostatic potential of the *Haloferax* CS, indicating, that 0.6M KCl is sufficient to screen the excess negative charge.

Figure 4.3 The Molecular Surface of Pyrococcus furiosus Citrate Synthase calculated in 0 M KCl



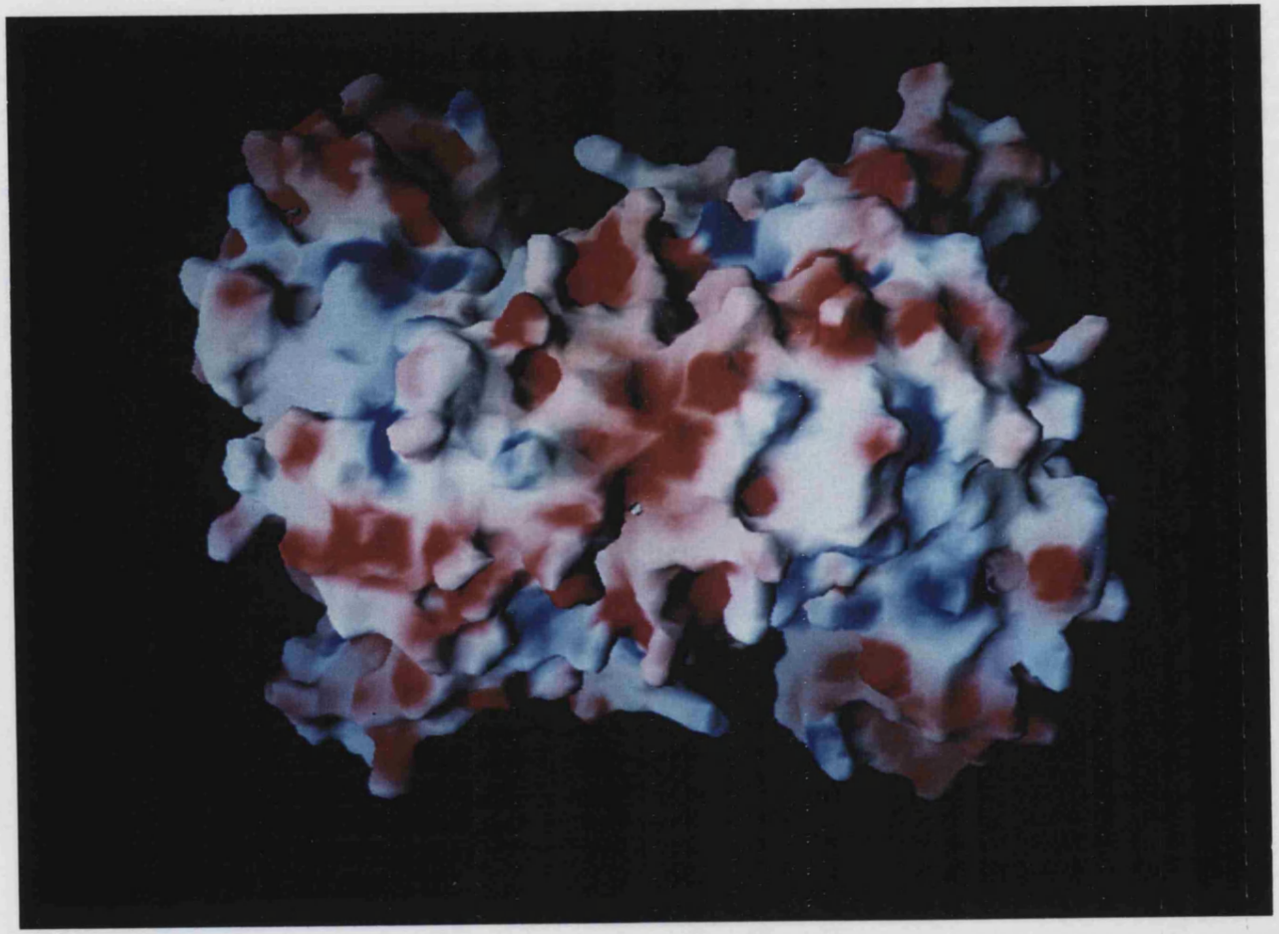
The molecular surface of citrate synthase from *Pyrococcus furiosus*. The surface profile of the citrate synthase was calculated and displayed using the programme GRASP. The surface is coloured according to the local electrostatic potential, calculated in 0 M KCl; red indicates negative potential and blue indicates positive potential. The structure is in the same orientation as the *Hf volcanii* citrate synthase modelled ribbon structure (Fig 4.2)

Figure 4.4 The Molecular Surface of the Model of Haloferax volcanii Citrate Synthase calculated in 0 M KCl



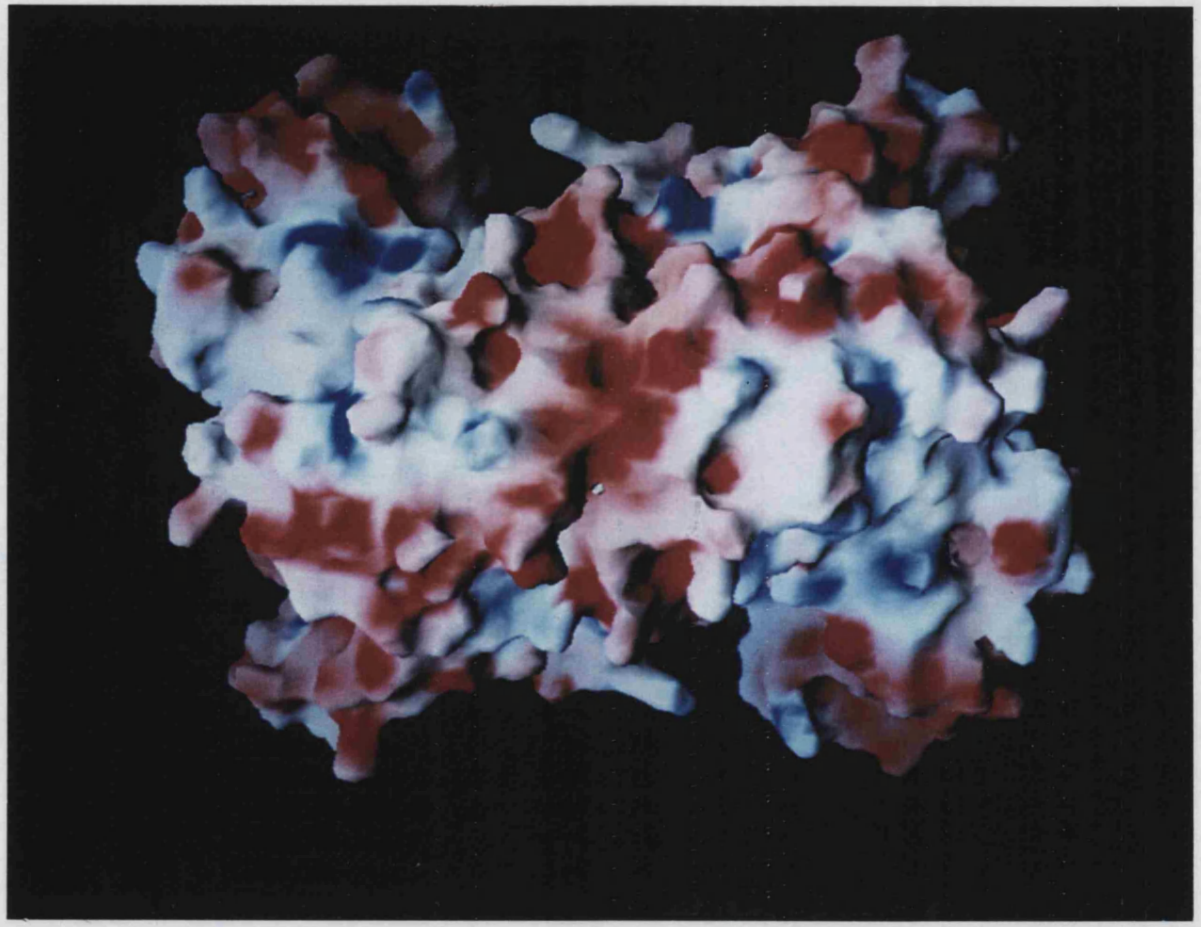
The molecular surface of the model citrate synthase from *Haloferax volcanii*. The model is based upon the high resolution crystal structure of *Pyrococcus furiosus*. The surface profile of the citrate synthase was calculated and displayed using the programme GRASP. The surface is coloured according to the local electrostatic potential, calculated in 0 M KCl; red indicates negative potential and blue indicates positive potential. The structure is in the same orientation as the *Hf volcanii* citrate synthase modelled ribbon structure (Fig 4.2)

Figure 4.5 The Molecular Surface of the Model of Haloferax volcanii Citrate Synthase calculated in 0.6 M KCl



The molecular surface of the model citrate synthase from *Haloferax volcanii*. The model is based upon the high resolution crystal structure of *Pyrococcus furiosus*. The surface profile of the citrate synthase was calculated and displayed using the programme GRASP. The surface is coloured according to the local electrostatic potential, calculated in 0.6 M KCl; red indicates negative potential and blue indicates positive potential. The structure is in the same orientation as the *Hf volcanii* citrate synthase modelled ribbon structure (Fig 4.2)

Figure 4.6 The Molecular Surface of the Model of Haloferax volcanii Citrate Synthase calculated in 2 M KCl



The molecular surface of the model citrate synthase from *Haloferax volcanii*. The model is based upon the high resolution crystal structure of *Pyrococcus furiosus*. The surface profile of the citrate synthase was calculated and displayed using the programme GRASP. The surface is coloured according to the local electrostatic potential, calculated in 2 M KCl; red indicates negative potential and blue indicates positive potential. The structure is in the same orientation as the *Hf volcanii* citrate synthase modelled ribbon structure (Fig 4.2)

5.3.1 Comparison of Surface Profiles of Citrate Synthase Structures

A comparison of surface profiles of other known CS structures; psychrophilic isolate DS2-3R, Pig, *Thermoplasma* and *Sulfolobus* can be viewed in simulated low salt conditions (0M KCl), in context with the *Haloferax* and *Pyrococcus* surface profiles already shown (Figure 5.7). The structures are displayed to give views across the dimer interface to reveal the active site. Most structures, with the exception of *Haloferax*, show the predicted profile of a typical hydrated protein with a mixture of negative and positive charges interspersed with neutral regions. There is not surprisingly a higher concentration of positive charges around the active site region due to the CS substrates being negatively charged. The increase in positive residues around the active site region is shown in stark contrast to the excess negatively charged residues present on the surface of the *Haloferax* modelled enzyme.

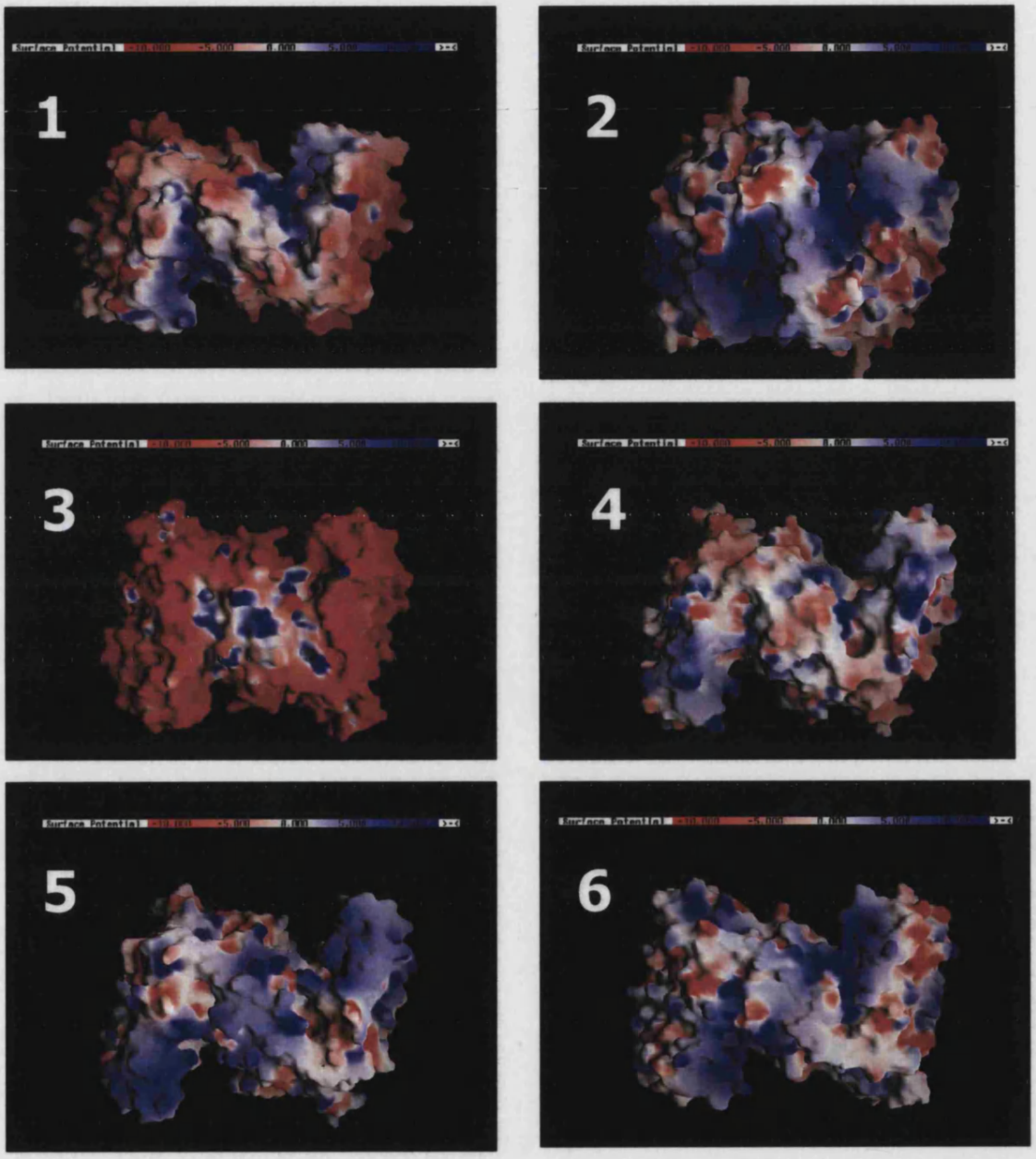


Figure 5.7
Comparative Views of the Surfaces of the Halophilic, Mesophilic, Psychrophilic, and Thermophilic CSs to Illustrate the Different Surface Properties.

1-6, views across the dimer interface to show the active site of the DS2-3R, Pig, *Haloferax Thermoplasma*, *Sulfolobus* and *Pyrococcus* enzymes respectively, to show the electrostatic potential at 0M salt concentration, prepared using the program GRASP. Red corresponds to a surface potential of less than -10 kcal (mol.electron) $^{-1}$ and blue corresponds to a potential of greater than +10 kcal (mol.electron) $^{-1}$

Table 5.1 Comparison of the Haloferax CS Model with the Most Negatively Charged Water-Soluble Proteins in the Protein Data Bank.

Protein	Source Organism	Net Charge	No. Residues	Charge Per Residue	Accessible Surface Area (Å^2)	Net Charge Density X 10^3 (Å^2)
*Pepsin	Pig	-39	326	-0.12	13592	-2.9
*Troponin C	Turkey	-30	162	-0.19	10001	-3.0
*Calmodulin	Drosophila	-23	148	-0.16	10289	-2.2
*Malate DH[1]	*H.marismortui*	-32	303	-0.11	22725	-1.4
*Ferredoxin	*H.marismortui*	-28	128	-0.22	6532	-4.3
#Citrate Synthase[2]	*H.volcanii*	-27	379	-0.07	14647	-1.8

The accessible surface area was calculated with program DSSP (Kabsch & Sander,1983).
All figures quoted are for the respective monomers.
[1] Proteins are tetramers in solution. [2] Proteins are dimers in solution.
*The information was taken from Table 2: The most negatively charged water-soluble proteins in the Protein Data Bank. (Frolow *et al.,* 1996).
#Data provided are based upon a gene homology model, not a crystal structure.

When comparing the model of the Haloferax CS with the most negatively charged proteins in the Protein Data Bank and with other halophilic crystal structures, HfvCS definitely exhibits halophilic traits, but does not stand out exceptionally amongst the other proteins in Table 5.1.

Interestingly, HmFd, considered to be a moderately halophilic enzyme, has the highest surface negative charge density of any protein structure in the Protein Data bank (Frolow *et al.*, 1996). HfvCS has the lowest charge/residue (-0-07e) of the proteins mentioned in Table 5.1, but a net charge density (-1.8 x 10^{-3}e Å^{-2}) comparable with its negatively charged counterparts. This is agreement with the findings of the simulated surface profile models (Figs 5.4-5.6) reiterating the high negative electrostatic potential located on the solvent accessible surface area.

Table 5.2 Comparison of HfvCS Model With Other CS Homologs Whose Crystal Structure Has Been Defined.

Source Organism	Net Charge	No. Residues/ Monomer	Charge/ Residue	Accessible Surface Area ($Å^2$)	Net Charge Density X 10^4 ($Å^2$)
DS2-3R	+28	380	+0.03	26614	+10.5
Pig	+26	437	+0.03	29158	+8.9
S.solfataricus	+11	378	+0.01	27501	+4.0
T.acidophilum	+13	384	+0.02	26643	+4.9
P.furiosus	+15	376	+0.02	26596	+5.6
*H.volcanii	-53	379	-0.07	29315	-18.1

The accessible surface area was calculated with program DSSP (Kabsch & Sander,1983).
All figures quoted are for the respective dimers.
*Data provided is based upon a molecular model, not a crystal structure.

Comparing the accessible electrostatic charges of the HfvCS model with other crystal structures of the same enzyme reveals some stark contrast (Table 2). Here, HfvCS does stand out as being clearly different in terms of net electrostatic charge, being the only protein to exhibit an overall negative potential and a larger surface area.

Table 5.3 Comparison of Amino Acid Residues, Present in Homologs of Citrate Synthase Whose Crystal Structures have been Resolved.

Amino Acid residues	DS2-3R	Pig	*T.acidophilum*	*S.solfataricus*	*P.furiosus*	*H.volcanii*
Nonpolar	51.0	51.3	50.8	49.6	51.1	50.1
Polar	22.9	25.2	22.1	22.5	20.2	19.8
Charged	26.1	23.5	27.1	27.9	28.7	30.1
Arg	19 (5)	19 (4.3)	19 (4.9)	19 (5)	17 (4.5)	21 (5.5)
His	11 (2.9)	14 (3.2)	6 (1.6)	5 (1.3)	8 (2.1)	8 (2.1)
Lys	18 (4.7)	25 (5.7)	30 (7.8)	32 (8.5)	33(8.8)	13 (3.4)
Glu	31 (8.2)	24 (5.5)	30 (7.8)	31 (8.2)	39 (10.4)	35 (9.2)
Asp	20 (5.3)	21 (4.8)	19 (4.9)	18 (4.8)	11 (2.9)	37 (9.8)

Numbers were generated using PEPTIDESORT (GCG).
Figure representing amino acid groups are defined as percentages.
Individual amino acid residue figures are absolute, those in brackets are representative percentages.

When comparing amino acid groups within CS across the species represented (Table 5.3), HfvCS does have a slightly higher percentage of charged residues, however the difference is not statistically significant. What is significant is the distribution of charged residues.

Halophilic proteins are widely known to have an excess of negatively charged residues, in particular glutamate, compared to their non-halophilic homologs. It has been suggested that glutamate plays a role in binding water molecules due to the presence of carboxylate groups (Kuntz, 1971). However, glutamate residues are not seen to be statistically in abundance in the *Haloferax* protein, with the *Pyrococcus* enzyme even having a slightly higher concentration. It is the aspartate residues that appear to be significantly more abundant in the *Haloferax* CS, being at least twice the number of these residues than in any of the other non-halophilic homologs.

It has been proposed that salt bridges, especially inter-monomer attractions may contribute to thermostability, as has been observed in the crystal structure of CS from the hyperthermophile *P.furiosus* (Russell *et al.*, 1997). Similar observations have also been made for the crystal structure of the extremely halophilic enzyme HmMDH (Dym *et al.*, 1995), as regards halophilic stability. The calculated number of ion-pairs of the HfvCS model has been compared with two thermophilic, and a mesophilic CS crystal structures at three different cut-off distances. There appears to be no significant increase in the number of salt-bridges present, when compared with either the thermophilic enzyme, (which shows an increased number of ion-pairs compared to the mesophilic structure), or the mesophilic structure. HfvCS appears, if anything, to have a significantly lower number of ion-pairs present than any of the other structures mentioned in the table. The inference from Table 5.4 is that salt-bridges do not appear to play a vital role in conferring stability regarding haloadaptation, although caution must be adhered to, as the predictions made are based upon a molecular model.

5.4 Biological Implications

HfvCS is active and stable at KCl concentrations as low as 0.5M; therefore it must be classed as a moderately halophilic enzyme. Other structures of moderately halophilic enzymes include HmFd (Frolow *et al.*, 1996), and DHFR (Pieper *et al.*, 1997) , both of which remain active and stable at low monovalent salt concentrations (~0.5M). HmMDH is considered to be extremely halophilic losing its activity below 2.4M NaCl (Eisenberg *et al.*, 1992).

The HfvCS clearly has the high acidic amino acid residue content associated with halophilic proteins. The proposed model structure of the enzyme places the majority of these

Table 5.4 Ion Pairs in Citrate Synthase[1]

	Pig	*T.acidophilum*	*P.furiosus*	*H.volcanii*
		6 Å cutoff		
Total ion pairs	82 (315)	90 (318)	105 (376)	70 (256)
Total ion pairs (A)	35 (124)	39 (132)	42 (119)	28 (91)
Intra ion pairs (B)	35 (124)	41 (137)	39 (115)	28 (91)
Inter A to B	12 (67)	10 (48)	24 (142)	14 (74)
		5 Å cutoff		
Total ion pairs	64 (271)	65 (252)	69 (294)	46 (194)
Total ion pairs (A)	26 (102)	30 (109)	25 (82)	17 (63)
Intra ion pairs (B)	26 (102)	29 (107)	24 (83)	17 (63)
Inter A to B	12 (67)	6 (35)	20 (128)	12 (69)
		4 Å cutoff		
Total ion pairs	36 (185)	43 (184)	43 (217)	19 (102)
Total ion pairs (A)	12 (59)	21 (81)	14 (50)	5 (22)
Intra ion pairs (B)	12 (59)	18 (76)	12 (47)	5 (22)
Inter A to B	12 (67)	4 (27)	17 (119)	9 (58)
		4 Å cutoff		
Arg	22/38 (58%)	19/34 (56%)	24/33 (73%)	30/46 (65%)
Lys	18/50 (36%)	30/54 (56%)	29/64 (45%)	6/26 (23%)
His	10/28 (36%)	8/12 (67%)	6/16 (38%)	6/16 (38%)
Glu	28/48 (58%)	29/50 (58%)	43/76 (57%)	22/68 (32%)
Asp	20/42 (48%)	16/38 (42%)	9/22 (41%)	16/74 (22%)

[1]The number of ion pairs is given using at 6, 5, and 4 Å cutoff distance between charged groups.
An estimate of the Coulombic energy was calculated as $332.06/(Dr_{ij})$, in kcal/mol, where D is the distance of the centre of the ion pair from the centre of the mass of the dimer, in Å, and r_{ij} is the distance between the ions, in Å.
The percentage involvement in ion pairs of charged amino acids is calculated using the 5 Å cut-off distance.
A and B refer to the two monomers of the native dimer.

residues on the solvent-accessible surface, in keeping with other models and structures of halophilic proteins. The apparent requirement for such a high concentration of acidic residues, glutamate in particular, on the surface of halophilic proteins has been rationalised in terms of their superior ability to bind water molecules. Frolow, (1996) has proposed that these residues assist in stability by allowing the protein to compete with salt ions for free water at high salt concentrations, this ultimately achieved by the enzyme maintaining a hydration shell (Zaccai *et al.*, 1989) and preventing the protein from aggregation. Studies in keeping with this hypothesis have shown extremely halophilic proteins HmMDH (Boneté *et al.*, 1993) and EF-Tu (Ebel *et al.*, 1992) to bind more salt than their non-halophilic homologs.

DHFR has clusters of negative charges on the surface of the molecule which are only implicated in halophilicity because they are energetically repulsive at physiological pH, unless counteracted by cation binding. This reinforces the interpretation of the role of acidic residues previously put forward by Böhm & Jaenicke (1994) based on a homology model study of DHFR, suggesting that such an excess of negative charge can only be tolerated at high salt concentrations, and that in the absence of such a shielding effect, the charges would repel each other electrostatically.

Repulsive electrostatic interactions between acidic residues have been seen to be a major destabilizing factor in halophilic proteins in low salt concentrations; they have been shown to remain destabilising even at high salt concentrations (Elcock & McCammon, 1998). Consequently, the role that acidic residues play, may be more to prevent aggregation rather than to make a positive contribution to protein stability.

In an analysis of the distribution of water molecules around well ordered amino acid residues in high resolution crystal structures, it has been revealed that arginine sidechains, have been shown to bind more water molecules than lysine sidechains (Thanki *et al.*, 1988). Consequently, an increased arginine to lysine ratio has been observed in some halophilic proteins, (Jaenicke, 1991), such as HmMDH (Cendrin *et al.*, 1993) and HmFd (Frolow *et al.*, 1996).

When comparing HfvCS with the non-halophilic forms of the enzyme, this phenomenon is not observed. There does not appear to be any significant difference in the number of Arg residues of the chosen proteins, but there is a noticeable decrease in the lysine content when compared to the thermophilic forms of the enzyme. The other non-thermophilic forms of the enzyme have a similar ratio of Arg:Lys to HfvCS.

A lower content of hydrophobic amino acids compared to non-halophilic homologs (Lanyi *et al.*, 1974), is another important structural feature reported to be involved in conferring halophilicity.

The observation made by Lanyi on the hydrophobic amino acid content has not strictly been observed in this case. *Haloferax* CS has approximately the same number of hydrophobic residues as *Bacillus subtilis* CS, to which it has the highest sequence identity (Fig 4.4). However, it does have considerably fewer hydrophobic residues than its closest archaeal relative *Pyrococcus*.

5.5 conclusions

The homology model complies with the generally-accepted characteristics of halophilic proteins, namely a high degree of negatively-charged residues found primarily on the solvent-accessible surface. These may allow the protein to remain hydrated and soluble by attracting hydrated salt ions away from the bulk solvent, and may play an additional role in preventing aggregation at high salt concentrations.

In addition to these roles it has been proposed that the acidic residues may also be involved in salt bridge formation, leading to stabilisation of the folded native enzyme (Dym *et al.*, 1995). This observation has not been made in the case of HfvCS.

Reversed Lys:Arg ratios compared to non-halophilic structural homologs also implicated in conferring halophilicity in certain structures, has not strictly been observed in this particular case either.

The predictions made regarding halostability of HfvCS must be treated with a degree of caution, as they are based on an homology modelled structure. This underlies the need for the protein to be expressed, and resolvable crystals produced, to confirm the initial observations found here and to reinforce the current hypotheses on halophilicity.

Chapter 6

Expression of Citrate Synthase in its Native Host Haloferax volcanii, and in the Gram-Negative Bacterium Escherichia coli

6.1 Introduction

Now that the complete citrate synthase gene has been cloned and sequenced, it enables us to make structural predictions and form the basis for mutagenesis studies by means of a homology model (chapter 5). However, a major objective is to obtain the crystal structure of the enzyme and for this the protein needs to be expressed in sufficient quantities to enable crystallisation trials to be set up.

There are two approaches to recombinant protein expression, homologous expression in which the gene of interest is expressed in the native organism, and heterologous expression in which the protein is expressed in a foreign host.

6.2 Homologous Expression in Haloferax

Homologous expression of halophilic proteins has previously been documented by Jolley *et al.* (1996,1997) and Werber and Mevarech (1978).

6.2.1 Recombinant Expression using the pBAP5009 Construct

pBAP5009 is a *Haloferax/E.coli* shuttle vector that has been successfully used to express halophilic proteins in our research group (Fig 6.1). Its construction is based upon the plasmid pMDS24 (Jolley *et al.*, 1996), in which expression of the insert is under the control of the strong constitutive promoter from the rRNA operon of *Halobacterium cutirubrum* (Brown *et al.*, 1989). pBAP5009 differs only slightly from pMDS24 in that the mevinolin resistance marker has been changed for a novobiacin resistance marker. In the current work, pBAP5009 forms the basis of a new construct in which the citrate synthase gene is substituted for the dihydrolipoamide dehydrogenase (DHLipDH) gene (Fig 6.1).

6.2.2 PCR Amplification of the CS Gene to Incorporate Unique Restriction Sites

Primers were designed based upon the N and C terminal coding regions of the citrate synthase gene. Regions were required to incorporate both the start and stop codons, and to introduce the unique restriction sites *BamHI* and *KpnI* with minimal change to the amino acid sequence. These sites were introduced to enable the PCR product to be cloned into the *Haloferax* shuttle vector pBAP5009. The incorporation of the restriction sites into the primers is shown in Figure 6.2 and the binding positions of the primers within the gene sequence is displayed in Figure 6.3. The primer sequences are listed in chapter 2, Table 2.11.

Figure 6.1 The Haloferax Shuttle Vector for Use in Homologous Expression of Halophilic Proteins

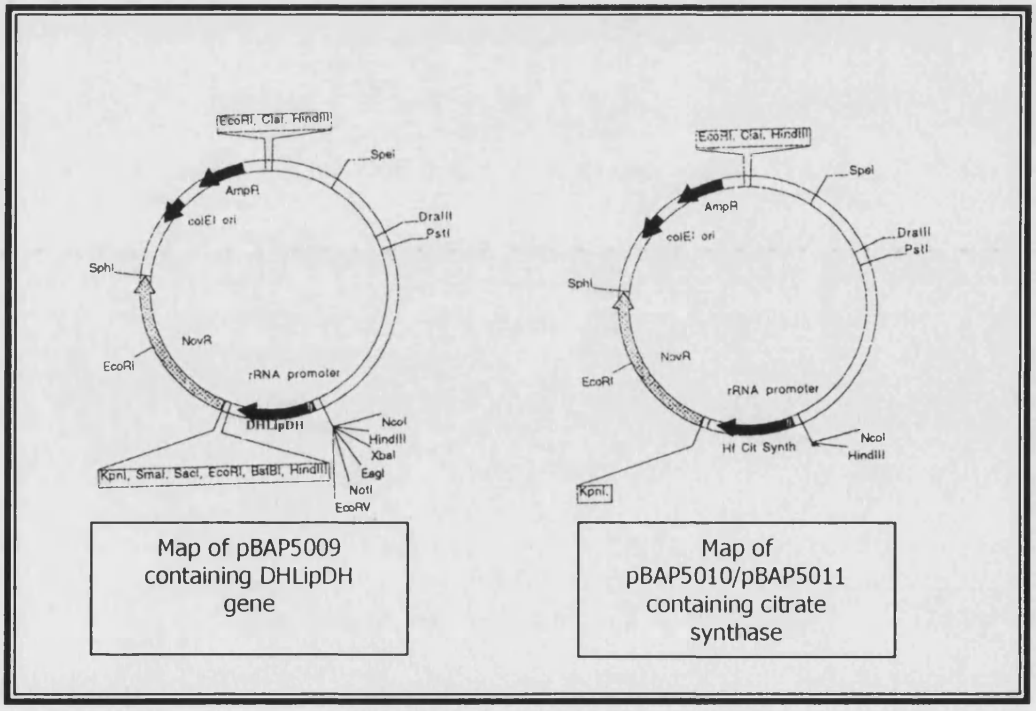
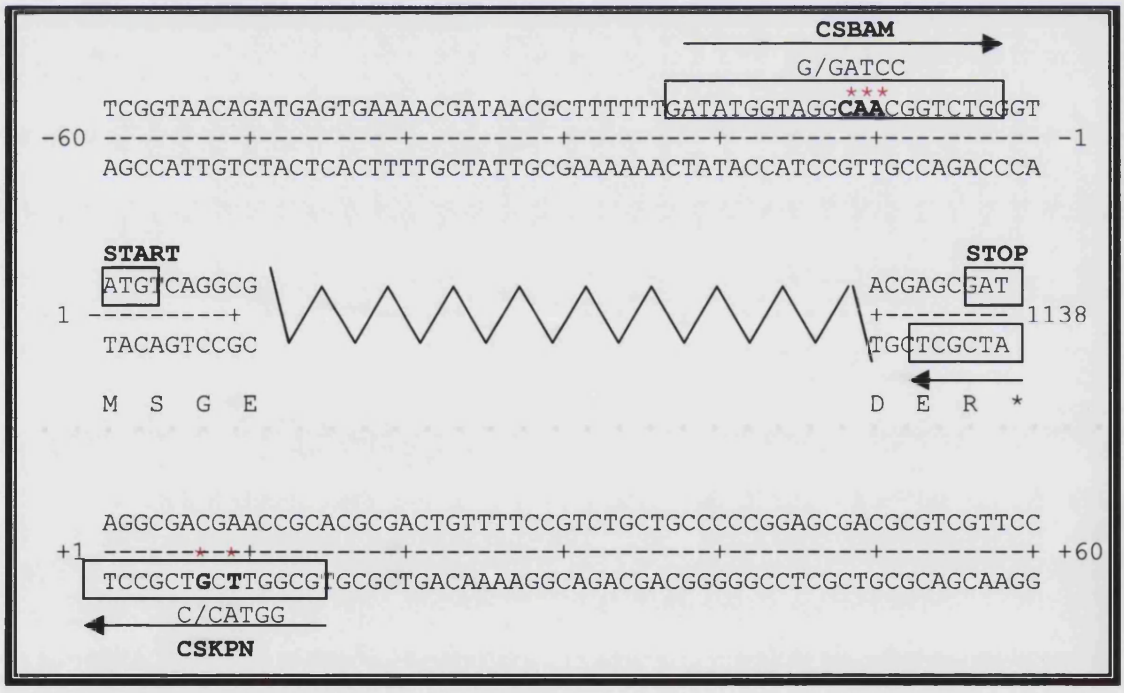


Figure 6.2 PCR Primers Used to Incorporate the Required Restriction Sites for Cloning into the Halophilic Expression Vector.

CSBAM:	5′ gAT ATg gTA **g/gA TCC** ggT CTg 3′	**21-mer**
CSKPN:	5′ Tg**C ggT AC/C** TCg CCT ATC gCT 3′	**21-mer**

* Indicates a nucleotide that has been altered from the wild type gene sequence to incorporate the appropriate restriction site. Restriction sites are embossed in red, / represents the cut site.

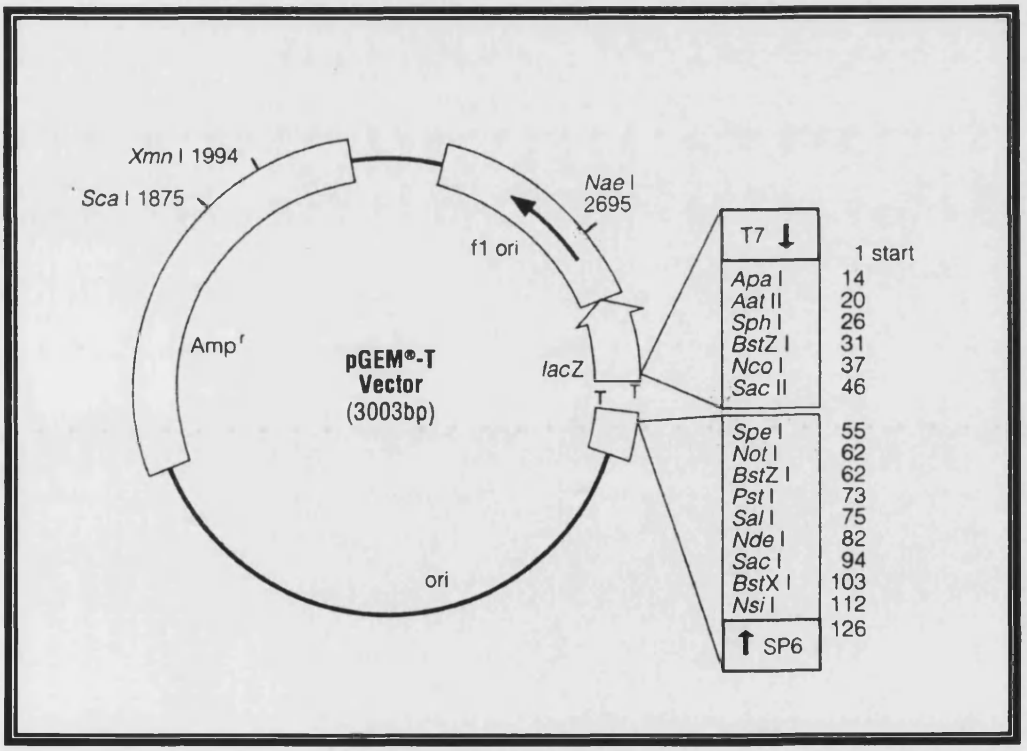
Figure 6.3 Binding Positions of the PCR Primers with Respect to the CS Gene Sequence.



Numbers refer to nucleotide positions with respect to the gene sequence.
Boxed areas indicate the sequence and binding positions of the primers.
* indicates a nucleotide that has been altered to incorporate the required restriction site.
Relevant restriction sites are shown above and below respective primer binding sites.
Arrows indicate direction of primer binding 5′ to 3′.

PCR amplification was conducted using the conditions and method outlined in chapter 2. A 1.2 kb product of the expected size was produced (Fig 6.5). The PRIMEZYME polymerase used appears to have been more specific, producing less artefacts. However, Vent DNA polymerase produced a higher concentration of the desired product. Both products were used in cloning reactions. The presence of the citrate synthase gene was initially confirmed by performing restriction digests on aliquots of the PCR products using enzymes chosen on the basis of the citrate synthase restriction map. PCR products were then cloned into the vector pGEMT® (Fig 6.4), so that the result of a digestion with *BamHI* and *KpnI* could clearly be seen. The clones were then restricted *with BamHI* and *KpnI* to release the citrate synthase fragment prior to its being isolated by electrophoresis and purified by the Geneclean® system.

Figure 6.4 The pGEM® -T Vector



Reproduced from the Promega Catalogue

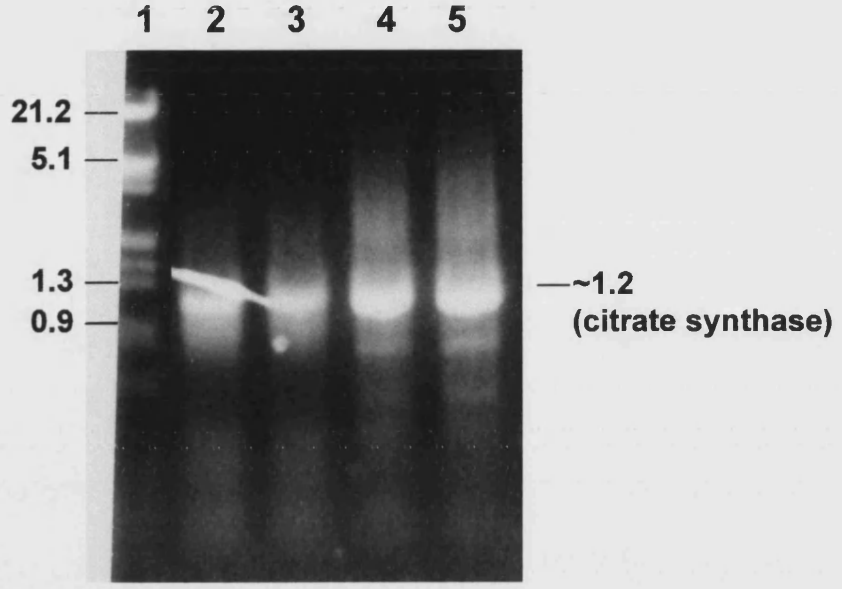


Figure 6.5
PCR Amplification of the Complete Citrate Synthase Gene using Primers with 'Unique' Restriction Sites Incorporated.

Lane 1: λ/H/E marker
Lane 2: 1.2 kb CS PCR product
Lane 3: 1.2 kb CS PCR product } **'PRIMEZYME' Taq.**
Lane 4: 1.2 kb CS PCR product
Lane 5: 1.2 kb CS PCR product } **Vent ® polymerase.**

The DNA was electrophoresed on a 1% (w/v) agarose gel. Fragment sizes are indicated in kilobases.

6.2.3 Ligation of CS into the pBAP5009 Shuttle Vector

The shuttle vector to be used in the cloning was only available as a recombinant clone (pBAP5009), containing the *Hf volcanii* DHLipDH gene (Jolley *et al.*, 1997). This gene had to be removed before the citrate synthase gene could be inserted into the vector.

A plasmid preparation of the shuttle vector pBAP5009 (1μg) was digested with the restriction enzymes BamHI and KpnI to remove the 1.6 kb DHLipDH insert. The products were then run out on a gel (Fig 6.6) and the 11 kb fragment corresponding to the shuttle vector alone was removed and purified using the Geneclean system®.

Ligations using rapid T4 DNA ligase were set up with the pBAP vector and citrate synthase insert by restriction of PCR products (6.2.2) and treated with shrimp alkaline phosphatase. Recombinant DNA was then transformed into *E.coli* (JM109) and plated onto agar plates supplemented with ampicillin (100μg/ml).

6.2.4 Positive PCR Screen for Recombinant Clones

23 potential recombinant colonies were visible after 12 h. Due to no instant recombinant selection mechanism being present in the plasmid apart from antibiotic resistance markers, it was decided to perform a PCR colony screen.

Using the primers CSBAM, used initially to introduce the unique restriction site for cloning, and DM9 used in the original sequencing of the λ clone (Chapter 3), a PCR screen was performed on the 23 colonies using part of the actual colony as a DNA source for the PCR reaction. λ clone 9 DNA was used as a positive control. 22 colonies gave positive results, showing a PCR product of the expected 700bp size (Fig 6.7). The colonies and PCR products were numbered according to the plate on which they grew.

Plasmid DNA was prepared from two colonies chosen at random (colony 14 and colony 23) using the Qiafilter® midi system (Fig 6.8). DNA was quantified by gel electrophoresis against a standard marker of a known concentration. The plasmids were digested with BamHI and KpnI to check for the presence of the citrate synthase insert. The original pGEMT® clone containing the citrate synthase gene was used as a control (Fig 6.9).

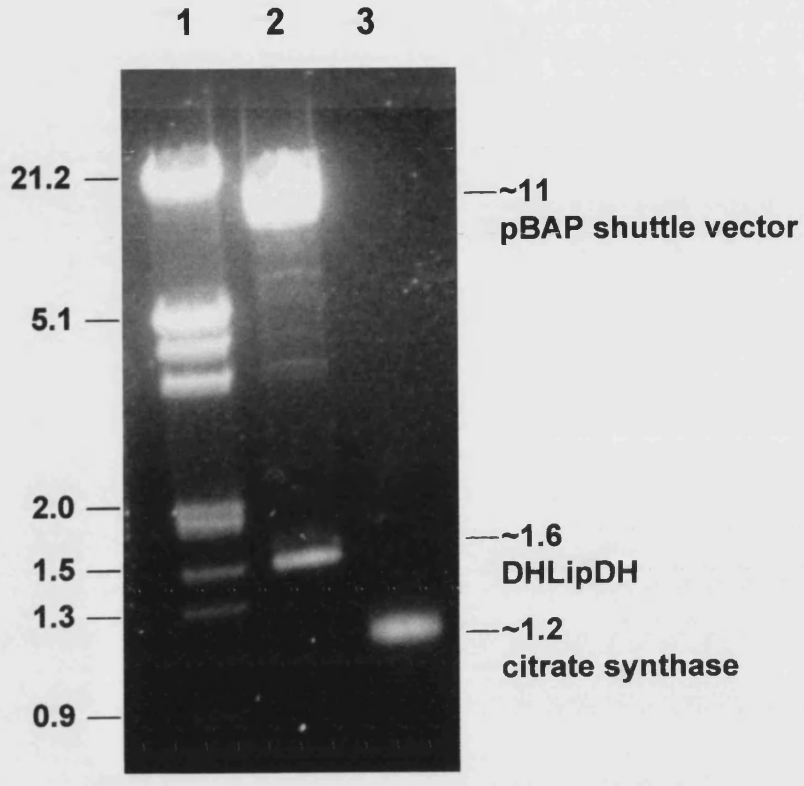


Figure 6.6
Digestion of Vector pBAP5009 to Remove DHLipDH Insert.

Lane 1: λ/H/E marker
Lane 2: pBAP5009 digested with *BamHI* & *KpnI*
Lane 3: 1.2 kb CS PCR product after digestion with *BamHI* & *KpnI*

The DNA was electrophoresed on a 1% (w/v) agarose gel. Fragment sizes are indicated in kilobases.

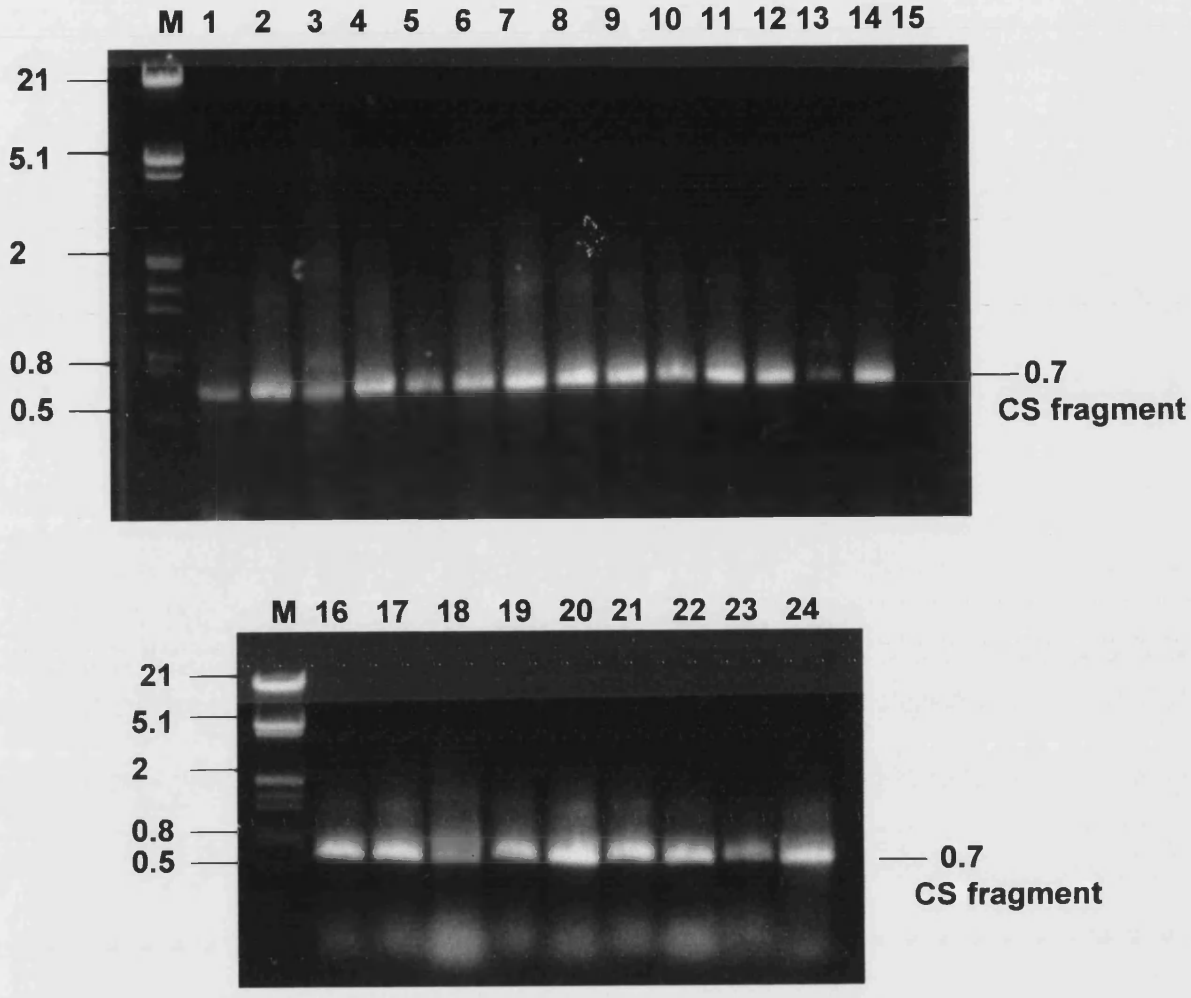


Figure 6.7
PCR (+ve) Screen for Recombinant pBAP5010 Clones.

Lane M: λ/Hind/Eco markers
Lanes 1-9: colonies from plate 1/ligation 1
Lanes 10-16: colonies from plate 2/ligation 2
Lanes 17-23: colonies from plate 3/ligation 3
Lane 24: control using λ clone 9 DNA

Fragment sizes are indicated in kilobases.
The PCR products were electrophoresed on 1% (w/v) agarose gels.

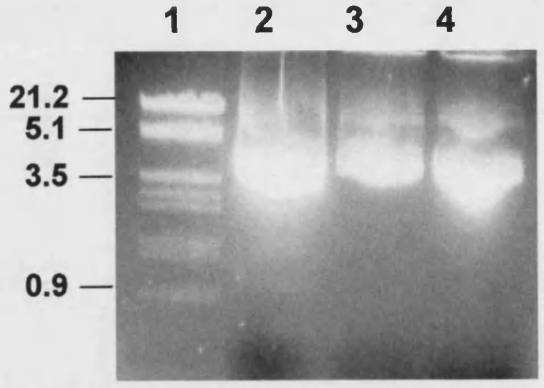


Figure 6.8
Analysis and Quantitation of Plasmid DNA prepared using Qiafilter® 100 Midi Preps.

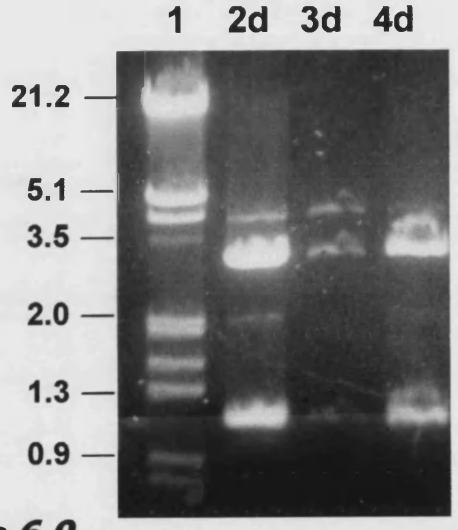


Figure 6.9
Digestion of Plasmid Preps (Fig 6.8) with BamHI & KpnI.

Lane 1: λ/H/E markers
Lane 2: colony 14
Lane 3: colony 23 } **See PCR screen (Fig 6.7)**
Lane 4: control- pGEMT® containing CS gene

The DNA was electrophoresed a 1% (w/v) agarose gel. Fragment sizes are indicated in kilobases.

6.2.5 Negative Screen for Recombinant Colonies

It appeared from the *BamHI* and *KpnI* digests of the chosen plasmids (Fig 6.9) that the two clones picked were the original pGEM®-T clones constructs containing the citrate synthase gene.

A second PCR screen was set up using M13 universal primers, the sequences of which can be found in chapter 2, Table 2.11; sequences complementary to the M13 primer sequence are present in the pGEM®-T vector but not in the pBAP shuttle vector. Here, a positive result indicated that the colony was not the desired transformant. DNA from 15 colonies (Table 6.1) gave positive results in the form of a 1.3kb PCR product (Fig 6.10). 10 ml cultures were grown from the colonies giving negative results (colonies 2-3,5-6,8-9,15 & 23), and plasmid minipreps were produced using the SNAP™ system (Fig 6.11), The plasmids were subsequently digested with *BamHI* and *KpnI* (Fig 6.12 & 6.13).

Table 6.1 Combined Results from Both PCR Screens

Screen No.	Primers	Expt size	No. +ve	No. -ve
I (Positive)	CSBAM DM9	700 bp	23 colonies No. 1-14, 16-23	1 colony No. 15
II (Negative)	M13 universal Forward/Reverse	1.3 kb	15 colonies 1,4,7,10-14,16-22	8 colonies 2-3,5-6,8-9,15,23

Colony numbers refer to the original colony grown and remain consistent for both screens.

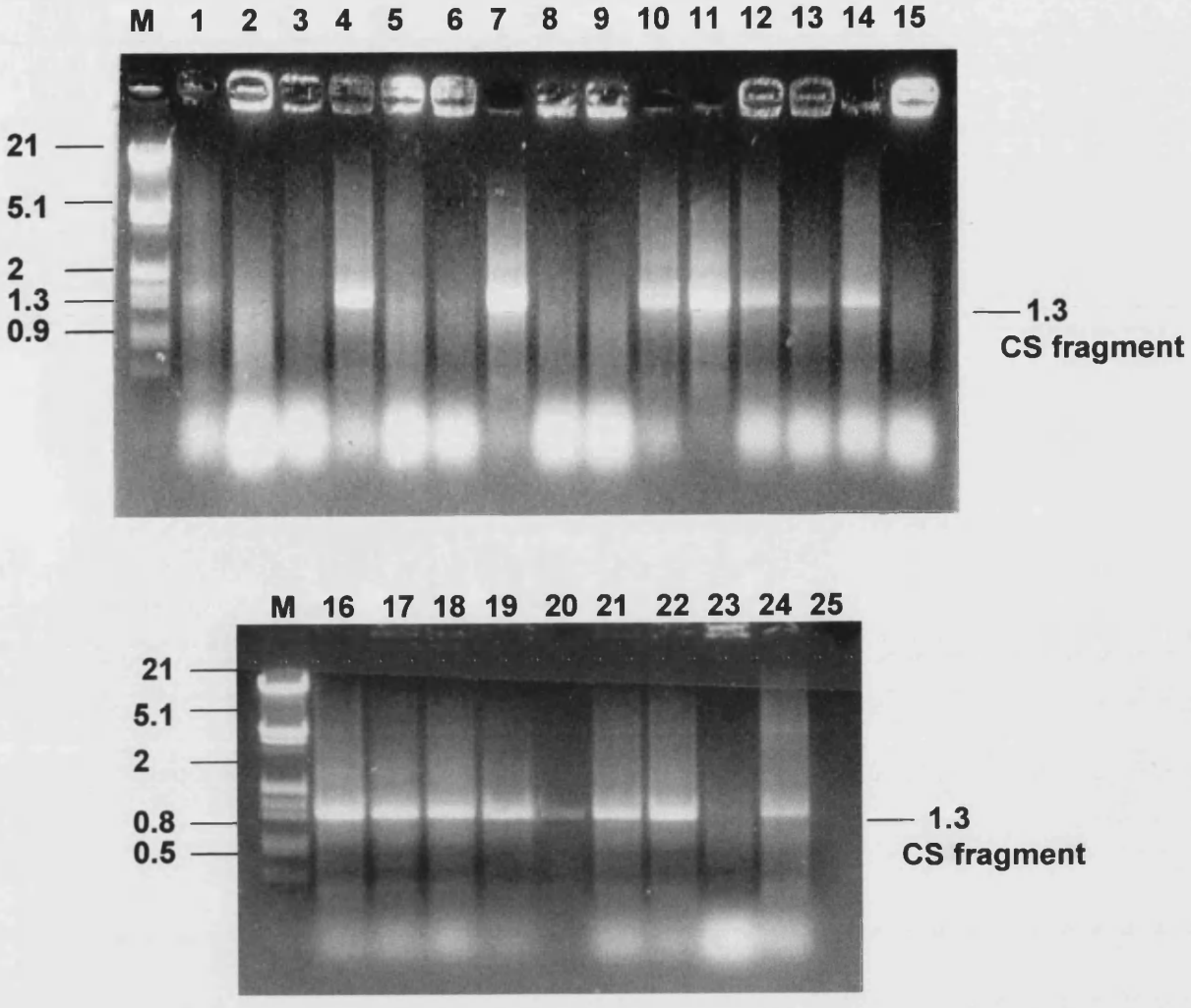


Figure 6.10
PCR (-ve) Screen for Recombinant pBAP5010 Clones.

Lane M: λ/Hind/Eco markers
Lanes 1-9: colonies from plate 1/ligation 1
Lanes 10-16: colonies from plate 2/ligation 2
Lanes 17-23: colonies from plate 3/ligation 3
Lane 24: +ve control using pGEMT®CS clone
Lane 25: -ve control using λ clone 9 DNA

Fragment sizes are indicated in kilobases.
The PCR products were electrophoresed on 1% (w/v) agarose gels.

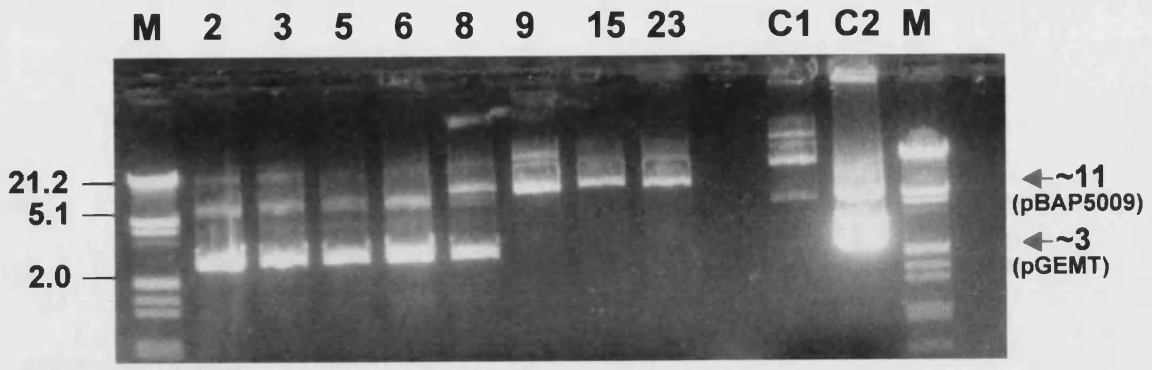


Figure 6.11
Plasmid Minipreps Performed with Regard to the Negative Results Obtained from PCR Screen (Fig 6.6).

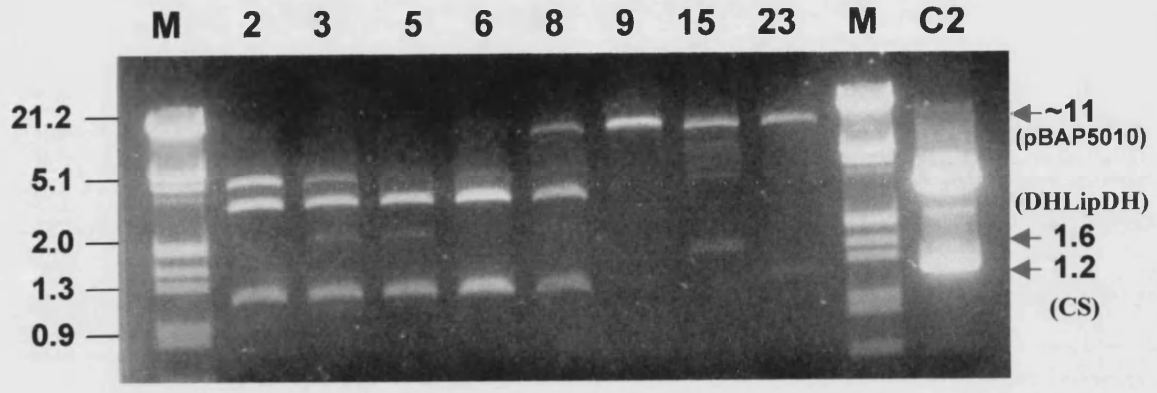
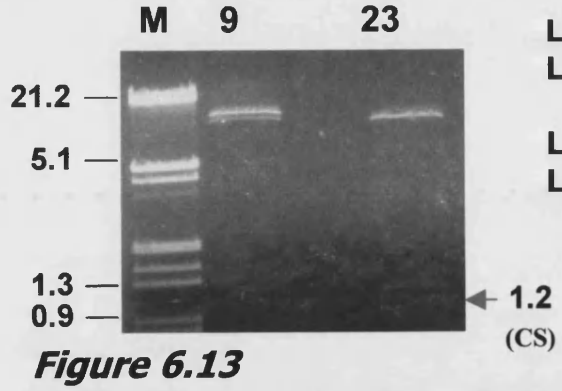


Figure 6.12
Digests of Above Plasmid Preps (Fig 6.11) Digested with BamHI & KpnI .



Lanes M: λ/H/E markers
Lanes No. 2-22: Refers to colony No. from PCR screen (Fig 6.6)
Lane C1: Refers to pBAP5009
Lane C2: Refers to pGEMT® CS clone

Figure 6.13
Scale-Up Digests (BamHI & KpnI) Performed on Potential pBAP5010 Clones.

The DNA was electrophoresed on 1% (w/v) agarose gels. Fragment sizes are indicated in kilobases.

Table 6.2 Results from Digests of Plasmid DNA that gave a Negative Result in the 2nd PCR Screen.

Colony No.	2	3	5	6	8	9	15	23
Vector size kb	3	3	3	3	3	11	11	11
Insert size kb	1.2	1.2	1.2	1.2	1.2	1.2	1.6	1.2

Colony numbers refer to the original colony grown and remain consistent.

From the above tables (6.1 & 6.2), it appears that colonies 1-8, 10-14 and 16-22 were all the original pGEM®-T construct containing the *Hf volcanii* citrate synthase gene. Colony 15 had a vector of the desired size, but it contained a 1.6 kb insert and was assumed to be the original pBAP construct pBAP5009 containing the DHLipDH gene. Colonies No. 9 and 23 possessed both a vector and insert of the correct size (11 kb) and (1.2 kb) respectively, and were taken to be the recombinant clones of interest (Fig 6.12). This was subsequently confirmed by sequencing. The recombinant plasmid construct was renamed pBAP5010.

Plasmid DNA of pBAP5010 from (colony 23) (Fig 6.14) was made in preparation for transforming into its native host *Haloferax volcanii* using Qiafilter® 500 maxiprep system. The plasmid was run on a gel to assess quality and quantity. pBAP5010 was resuspended at a concentration of 250 ng/μl. A restriction digest was performed on the plasmid to ensure that the 1.2 kb insert it contained was in fact the citrate synthase gene. Aliquots of the plasmid were digested in either single or double digests using *BamHI*, *KpnI* or *XhoI* (Fig 6.15). *BamHI* and *KpnI* are known to be unique restriction sites within the shuttle vector. All digests produced a fragment of the expected size *Bam/Kpn* 1.2 kb, *Bam/Xho* 0.4 kb and *Kpn/Xho* 0.8 kb, indicating that the fragment was the citrate synthase gene.

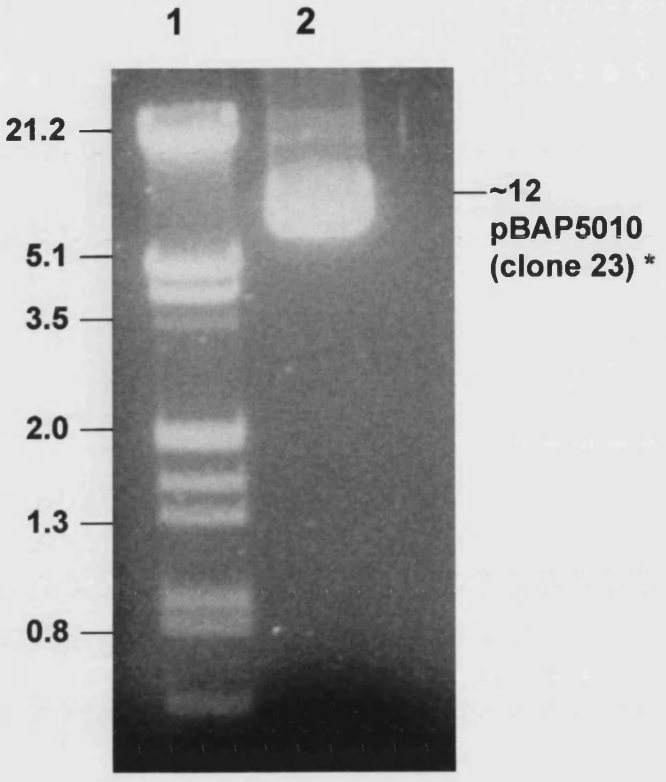


Figure 6.14
Analysis and Quantitation of pBAP5010 (Clone 23) Prepared Using Qiafilter® 500 Maxi Prep.

Lane 1: λ/H/E marker
Lane 2: pBAP5010 ~ 250 ng/μl.

The DNA was electrophoresed on a 1% (w/v) agarose gel.
Fragment sizes are indicated in kilobases.
*** Clone 23 refers to the colony No. see PCR screen (Fig 6.10).**

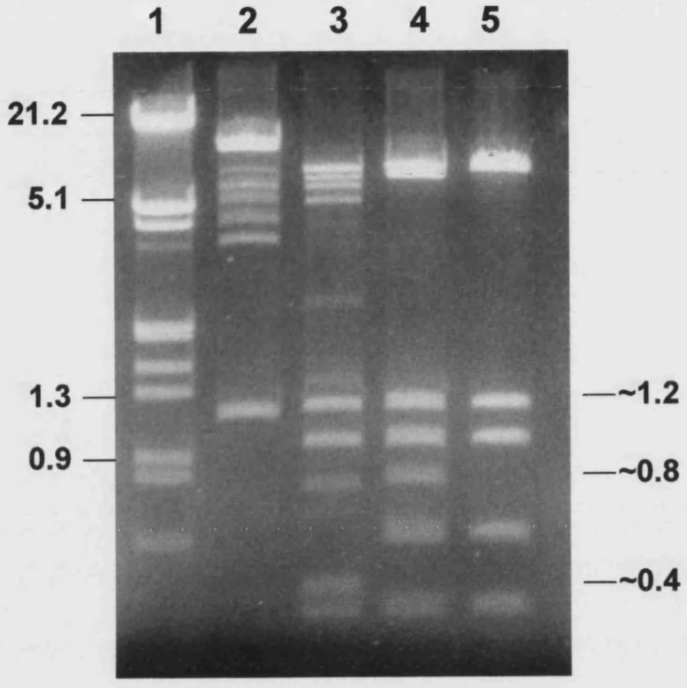


Figure 6.15
Restriction Digests Performed on pBAP5010 (Clone 23) to Ensure the Presence of CS.

Lane 1: λ/H/E marker
Lane 2: pBAP5010 digested with *BamHI/KpnI*
Lane 3: pBAP5010 digested with *BamHI/XhoI*
Lane 4: pBAP5010 digested with *KpnI/XhoI*
Lane 5: pBAP5010 digested with *XhoI*

The DNA was electrophoresed on a 1% (w/v) agarose gel. Fragment sizes are indicated in kilobases.

6.2.6 Transformation of Haloferax volcanii

Competent *Hf volcanii* cells were made using the protocol outlined by Dyall-Smith *et al.* (1995) (Chapter 2). The cells were transformed using 60% PEG 600 and 1 μg, 2μg and 5 μg amounts of plasmid DNA. Transformant cells were given 12 h at 37°C to recover before being plated out in top agar containing the selection antibiotic, novobiocin (300μg/ml). Colonies of transformants took 10 days to become visible to the naked eye, after which they were grown up in 18% MGM media in the presence of novobiocin. Untransformed *Hf volcanii* cells, which cannot grow in the presence of the antibiotic, were used as a control.

Plasmid minipreps were performed on the transformed colonies using the SNAP™ miniprep system, the principle of which is based on the alkaline lysis method. Plasmid DNA from these minipreps was used as a DNA source for a PCR screening reaction using CSBAM/CSKPN primers, to ensure the presence of the citrate synthase gene construct in the *Haloferax* transformants. The PCR reaction yielded products of the 1.2 kb expected size, thus confirming the presence of the CS gene in the recombinant plasmids.

6.2.7 Purification of Citrate Synthase

The recombinant *Haloferax* cell pellet was resuspended in 10 mM Tris buffer pH 8 containing 2M KCl and 1 mM EDTA to a final volume of 2ml. Cells were lysed by sonication and cell debris removed by centrifugation. The resultant cell extract was diluted 10 fold with 10 mM Tris buffer pH 8 containing 1 mM EDTA prior to purification by affinity chromatography on Matrex Gel Red A. The use of low salt conditions (0.2M KCl) during purification of the protein does not cause denaturation of the enzyme, a phenomenon observed with other halophilic proteins when exposed to diluted buffer (Eisenberg *et al.,* 1992).

Initial characterisation of the recombinant enzyme revealed that CS expression in the recombinant strain was double that of the wild type host. This was not deemed to be a high enough level of protein production to put forward into crystallisation trials and future expression and characterisation of mutant enzymes, therefore a method to increase protein expression was sought.

6.2.8 Construction of Shuttle Vector pBAP5011

A new construct was made using primers CSKPN, as used previously in the construction of pBAP5010, and a new primer CSBAMUS (Fig 6.16) positioned upstream of CSBAM (Fig 6.17); designed to incorporate what is thought to be the genes natural promoter again incorporating a unique *BAMHI* site enabling the construct to be cloned into the pBAP *Haloferax* shuttle vector.

Figure 6.16 PCR primer to Incorporate the Required Restriction Sites for Cloning into the Haiophilic Expression Vector.

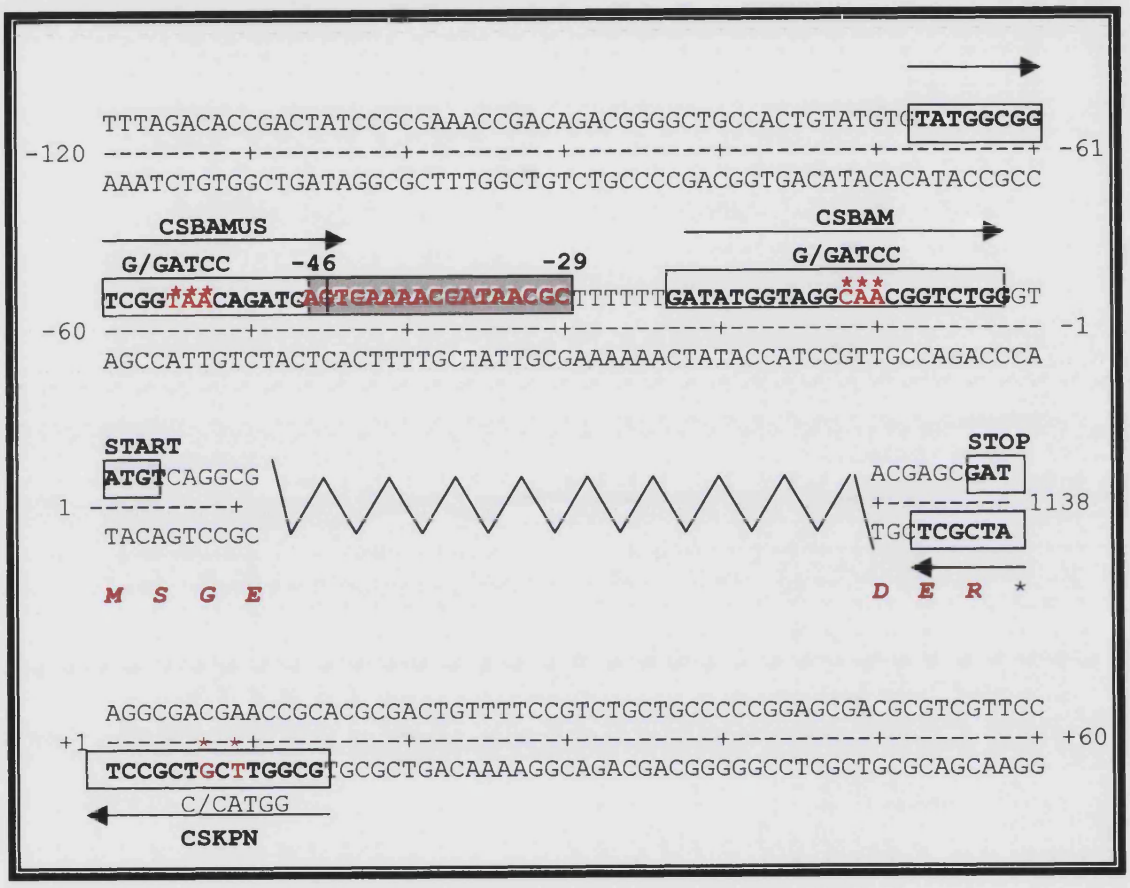
CSBAMUS: 5' TAT ggC ggT Cg/g ATC CAg ATg A 3' **22-mer**

* Indicates a nucleotide that has been altered from the wild type gene sequence to incorporate the appropriate restriction site. / Indicates the enzyme cut site.

The PCR was conducted using identical conditions to those used in the construction of pBAP5010. Two different makes of *Taq* polymerase was used in the PCR reactions, High Fidelity *Taq* and PRIMEZYME; PCR products from both enzymes appeared to be free from non-specific artefacts and were both seen to be of similar concentrations. Both PCR products were used in cloning reactions. 1.3kb fragments of the predicted size were seen when the PCR products were electrophoresed (Fig 6.18). The gene-cleaned products were cloned directly into the pBAP vector using the method outlined previously. The new sequence of the pBAP construct was confirmed and the plasmid was renamed pBAP5011.

Recombinant DNA was transformed into *E.coli* (JM109) and the cells were plated onto agar plates supplemented with ampicillin (100μg/ml). Potential recombinant colonies were screened using primers CSBAMUS and CSKPN (Fig 6.19). Qiagen maxipreps were performed to obtain sufficient quantities of DNA to transform *Haloferax*. *Haloferax* competent cells were prepared and transformed as described previously, and as was seen with the pBAP5010 transformants, it took 10 days for them to become visible to the naked eye.

Figure 6.17 Binding Positions of the PCR Primers with Respect to the Gene Sequence



Numbers refer to nucleotide positions with respect to the gene sequence.
Boxed areas indicate the sequence and binding positions of the primers.
* Indicates a nucleotide that has been altered to incorporate the required restriction site.
Relevant restriction sites are shown above and below respective primer binding sites.
Arrows indicate direction of primer binding 5′ to 3′.
☐ Indicates the position of the suspected haloarchaeal promoter.

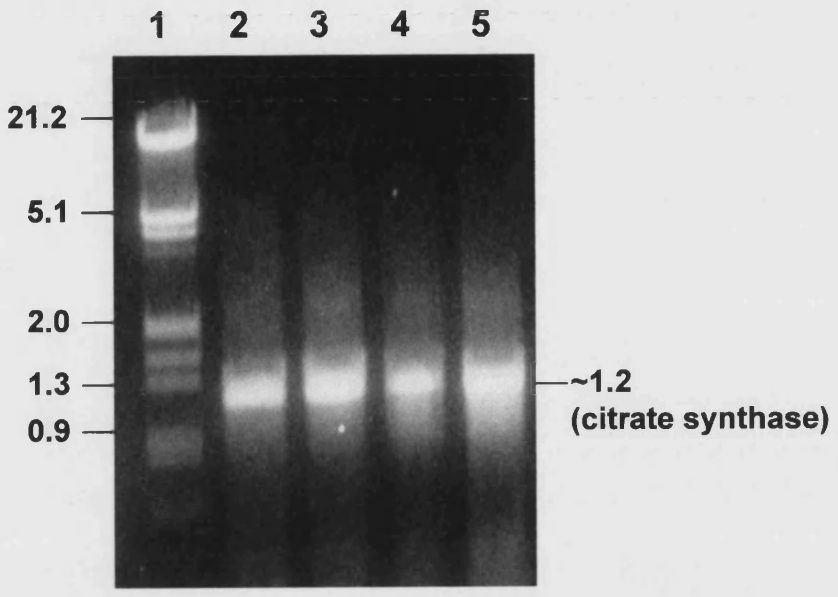


Figure 6.18
PCR of the Complete CS Gene with Proposed Upstream Promotor Region.

Lane 1: λ/H/E marker
Lane 2: 1.2 kb CS PCR product
Lane 3: 1.2 kb CS PCR product } ***High Fidelity Taq.***
Lane 4: 1.2 kb CS PCR product
Lane 5: 1.2 kb CS PCR product } ***'PRIMEZYME' Taq.***

The DNA was electrophoresed on a 1% (w/v) agarose gel. Fragment sizes are indicated in kilobases.

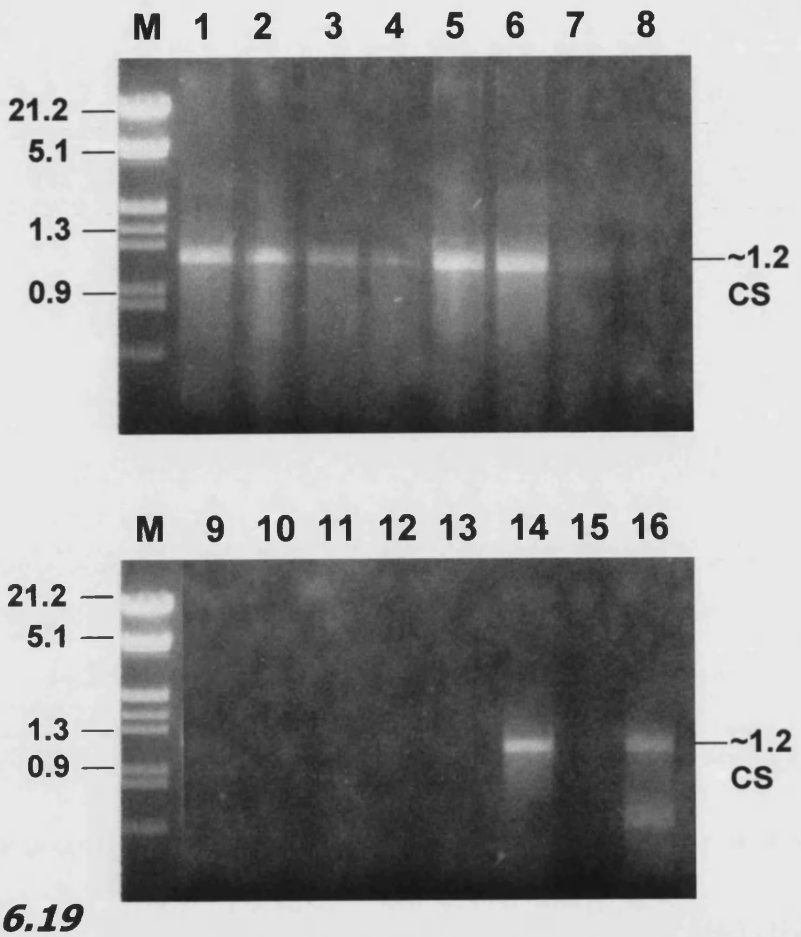


Figure 6.19
PCR Screen for Recombinant Shuttle Vector pBAP5011 Containing the Complete CS Gene and Proposed Upstream Promotor Region.

Lane M: λ/H/E markers
Lanes No. 1-16 refer to original colony number

The DNA was electrophoresed on 1% (w/v) agarose gels. Fragment sizes are indicated in kilobases.

6.2.9 Purification of the Recombinant Enzymes

Transformants were selected and grown in the appropriate volume of 18% MGM. The purification of CS was carried out from 100 ml of culture supplemented with novobiocin (0.3 μg/ml) using dye-affinity chromatography Matrex Red Gel A. Both recombinant and the native enzymes were extracted from *Haloferax* cells and purified at the same time. Fig 6.20- 6.22 show the SDS-PAGE gels of the purification procedure, Table 6.3 shows the yield and specific activities obtained.

Samples were assayed in 10 mM Tris (pH 8), 1 mM EDTA and 2M KCl at 30°C. The enzymes were purified to homogeneity: a single band with an M_r of 42 000 was visible when SDS-PAGE electrophoresis was performed (Fig 6.20-6.22).

Yields for the purification of the dimeric form of citrate synthase are normally in the region of 80-100% (Anderson *et al.*, 1994) when using Matrex Gel Red A (Amicon). However, Reactive Red (Sigma) was used in this instance; although this is supposed to be identical, practical results indicate that this may have a slightly different pH optimum (N.Heyer personal communication). This would account for the lower percentage yields.

Table 6.3 Purification Table of Native and Recombinant Homologousiy Expressed Citrate Synthase.

	Purification Step	Total Act (Units)	Total Protein (mg)	Spec. Act (Units/mg)	Yield (%)
Native	Cell Extract	3.3	22.0	0.15	100
	*Matrex Red A	1.0	.02	65	30
PBAP5010	Cell Extract	7.0	24.0	0.3	100
	Matrex Red A	4.6	.07	64	66
PBAP5011	Cell Extract	7.8	13.0	0.6	100
	Matrex Red A	4.0	.06	67	51

*Dye-affinity matrix used Reactive Red (Sigma)

Recombinant citrate synthase expressed using the pBAP5010 construct produced a two fold level of expression with respect to the wild type *Haloferax*, citrate synthase expressed from the pBAP5011 construct produced a 4.5 fold level of expression compared to the untransformed *Haloferax*.

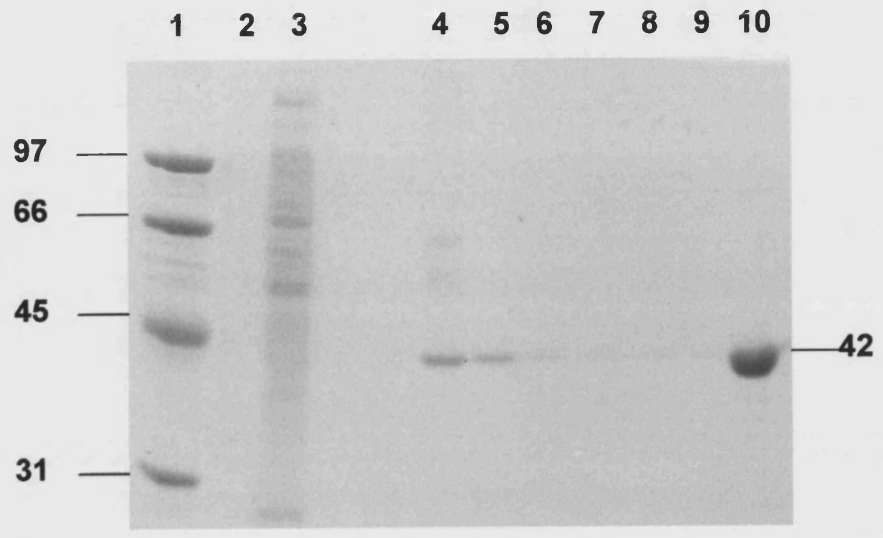


Figure 6.20
Purification of Citrate Synthase from Untransformed Hf. volcanii using Matrex Red Gel A Affinity Chromatography.

Lane 1: Low molecular weight markers
Lane 2: Insoluble debris after centrifugation of cell extract
Lane 3: Cell extract supernatent
Lane 4 -9: Fractions collected after Red Gel A purification
Lane 10: *Hf. volcanii* CS standard

Protein sizes are indicated in kilodaltons.
Proteins were electrophoresed on a 10% SDS polyacrlyamide gel.

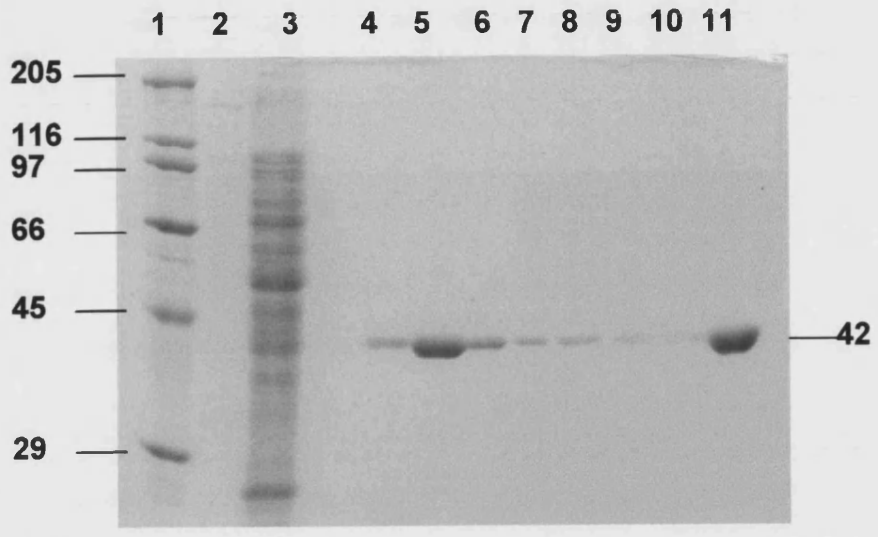


Figure 6.21
Purification of Recombinant Citrate Synthase from Hf. volcanii Transformed with pBAP5010.

Lane 1: Broad range molecular weight markers
Lane 2: Insoluble debris after centrifugation of cell extract
Lane 3: Cell extract supernatent
Lane 4 -10: Fractions collected after Red Gel A purification
Lane 11: *Hf. volcanii* CS standard

Protein sizes are indicated in kilodaltons.
Proteins were electrophoresed on a 10% SDS polyacrlyamide gel.

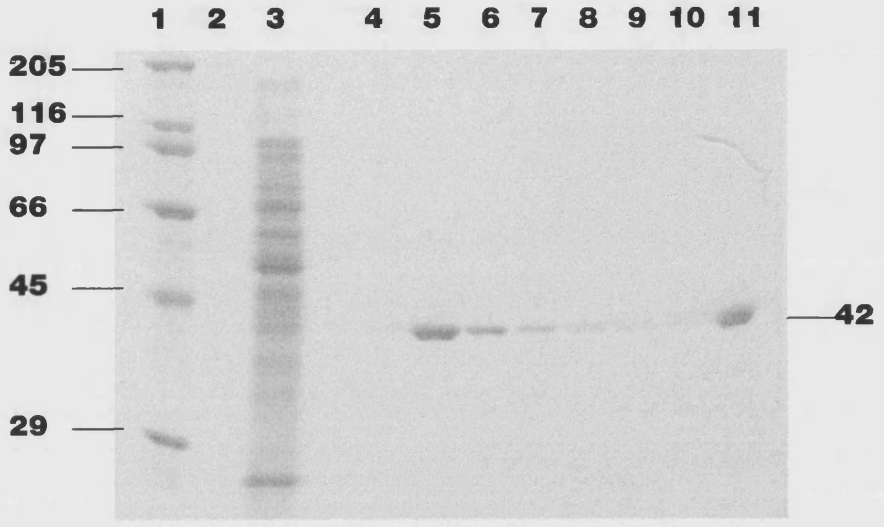


Figure 6.22
Purification of Recombinant Citrate Synthase from Hf. volcanii Transformed with pBAP5011.

Lane 1: Broad range molecular weight markers
Lane 2: Insoluble debris after centrifugation of cell extract
Lane 3: Cell extract supernatent
Lane 4 -10: Fractions collected after Red Gel A purification
Lane 11: *Hf. volcanii* CS standard

Protein sizes are indicated in kilodaltons.
Proteins were electrophoresed on a 10% SDS polyacrlyamide gel.

6.2.10 Kinetic Characterisation of Expression Products

Kinetic parameters were determined from assays of CS activity in 10 mM Tris (pH8) containing 1 mM EDTA, 2M KCl and purified enzymes. Assays were conducted at 30°C. The direct linear plot of Eisenthal and Cornish-Bowden (1974) was used to determine K_m and V_{max} values.

Table 6.4 Kinetic Parameters of Recombinant and Native Halophilic Citrate Synthase

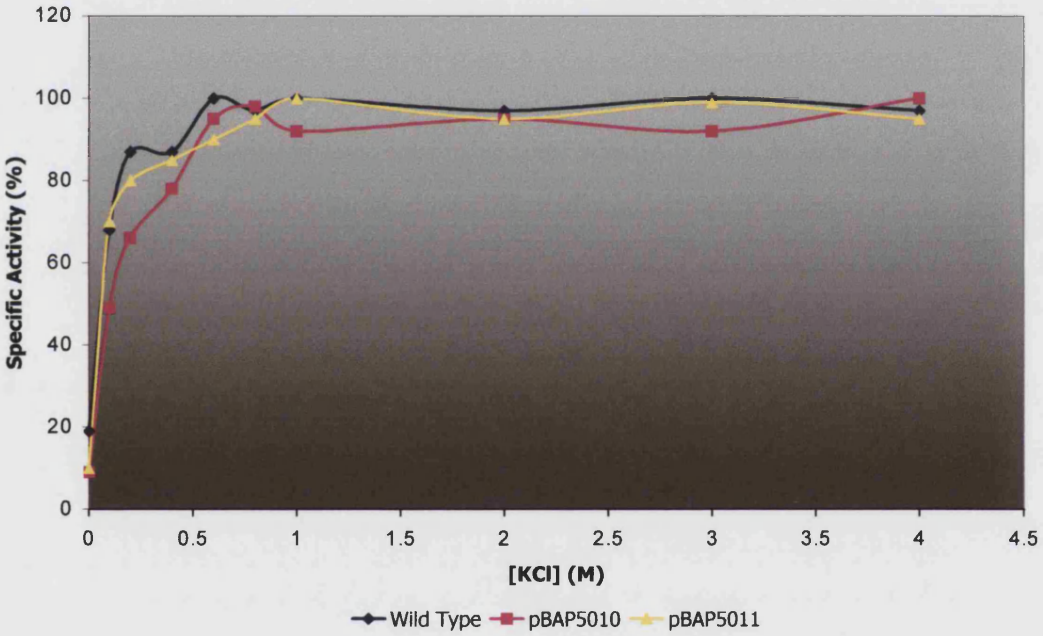
	Km (OAA) μM	Km (AcCoA) μM	V_{max} U/mg
Native	31 ± 2	148 ± 9.6	56 ± 0.9
PBAP5010	36 ± 2	212 ± 8.8	126 ± 3.5
PBAP5011	35 ± 2	199 ± 10.6	204 ± 3

Both recombinant enzymes have similar kinetic properties to the native enzyme, implying that no changes have taken place which significantly effect the activity of the enzyme due to recombinant expression in the pBAP vector system.The differences in the V_{max} values probably indicates that the native and pBAP5010 and pBAP5011 were not purified to homogeneity.

6.2.1.1 Dependency of Catalytic Activity on KCl Concentration

KCl profiles were determined in 1mM EDTA, 10 mM Tris (PH8), in the presence of varying concentrations of KCl ranging from 0–4M (Fig 6.23).

Figure 6.23 KCl Profiles of the Recombinant and Native Enzymes



In each case activity is expressed as a percentage of the maximum activity obtained.

As was seen with the kinetic parameters, both the native and recombinant enzymes showed very similar KCl profiles. 0.6-1.0 M proved to be a sufficient concentration of KCl for optimal activity of the enzyme. This is an interesting observation as 0.6M KCl was the concentration predicted by the programme GRASP to be sufficient to shield the high negative surface charge of the halophilic enzyme (Fig 5.5).

6.2.12 Conclusion

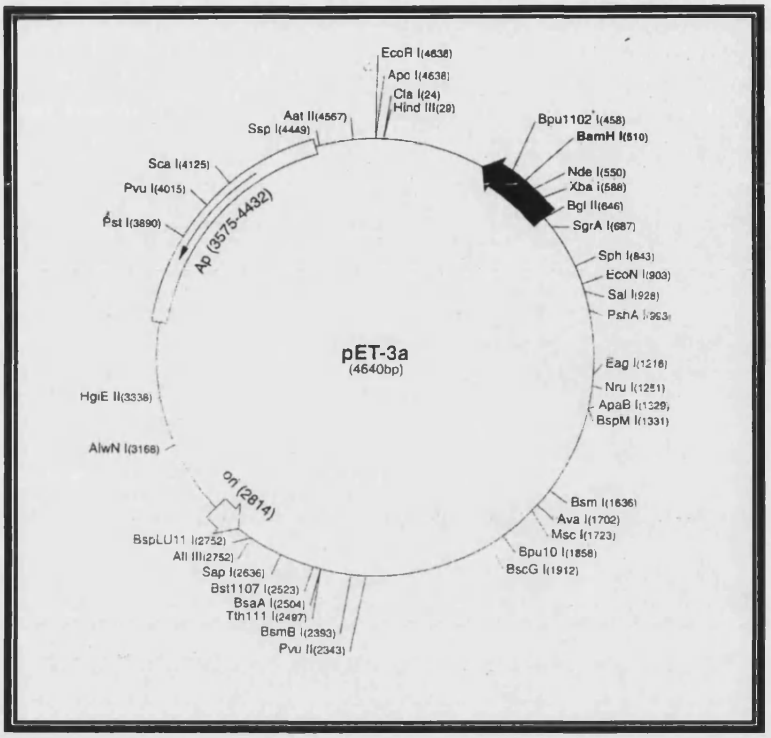
This quantity of protein expressed was not considered to be high enough for making use of the system as a way of a) producing protein for crystallisation trials and b) looking at the effects of site directed mutagenesis on halophilicity. Low yields of protein coupled with the slow rate of growth of *Haloferax* and the lack of a CS minus *Haloferax host* did not rate this system highly, and so an alternative expression system was sought.

6.3 Heterologous Expression in E.coli

Despite the potential problems of expressing a halophilic protein in a non-halophilic organism, such as refolding difficulties in a low salt environment, most of the halophilic protein structures that have successfully been resolved have in fact been determined from proteins expressed in *E.coli*. Halophilic genes successfully expressed in *E.coli* to date include malate dehydrogenase from *Halarcula marismortui* (Cendrin *et al.*, 1993), dihydrofolate reductase from *Hf.volcanii* (Blecher *et al.*, 1993), formylmethanofuran: tetrahydromethanopterin formyltransferase from *Methanopyrus kandleri* (Shima *et al.*, 1995), HMG-CoA from *Hf.volcanii* (Bischoff & Rodwell, 1996) and seryl-tRNA synthetase from *Haloarcula marismortui* (Taupin *et al.*, 1997).

As part of a collaboration with Dr. J. Chaudhuri [*Dept of Chemical Engineering, University of Bath*] work had been conducted simultaneously looking at alternative expression systems for halophilic proteins. Use of the pET expression system, which is under the control of strong bacteriophage T7 transcription (and translation) signals (Studier *et al* ., 1990) was investigated. Expression of the *Hf.volcanii* citrate synthase gene was achieved using the vector pET-3A in *E.coli* BL21 (DE3) (Connaris *et al.*, 1999). pET3A (Fig 6.24) is an inducible expression vector under the control of IPTG. Expression conditions were optimised to give recombinant citrate synthase with a specific activity in unfractionated cell exracts of 9 Units/mg. The recombinant enzyme was purified using Matrex Gel Red A (Amicon) dye-affinity chromatography as described previously.

Figure 6.24 The pET3A expression vector



Reproduced from the *Novogene* catalogue

As with homologously expressed proteins, host cells are lysed in the presence of 2M KCl and diluted the extract 10 fold for subsequent purification. Unlike homologously expressed proteins in *Haloferax*, *Haloferax* CS expressed in *E.coli* is not active upon lysing the cells, and full activity is only obtained after being exposed to 2M KCl for 12 h. The specific activity of the purified recombinant citrate synthase was calculated to be 40 Units/mg, and the enzyme was seen to be homogeneous with an apparent Mr of 42,000. Kinetic parameters were shown to be very similar both for the native and recombinant enzymes respectively.

The expression of recombinant citrate synthase in the *E.coli* host system had been undertaken in collaboration with Dr Helen Connaris and the Dept of Chemical Engineering, University of Bath.

6.4 Discussion

6.4.1 Construction of the Recombinant Plasmids pBAP5010 and pBAP5011

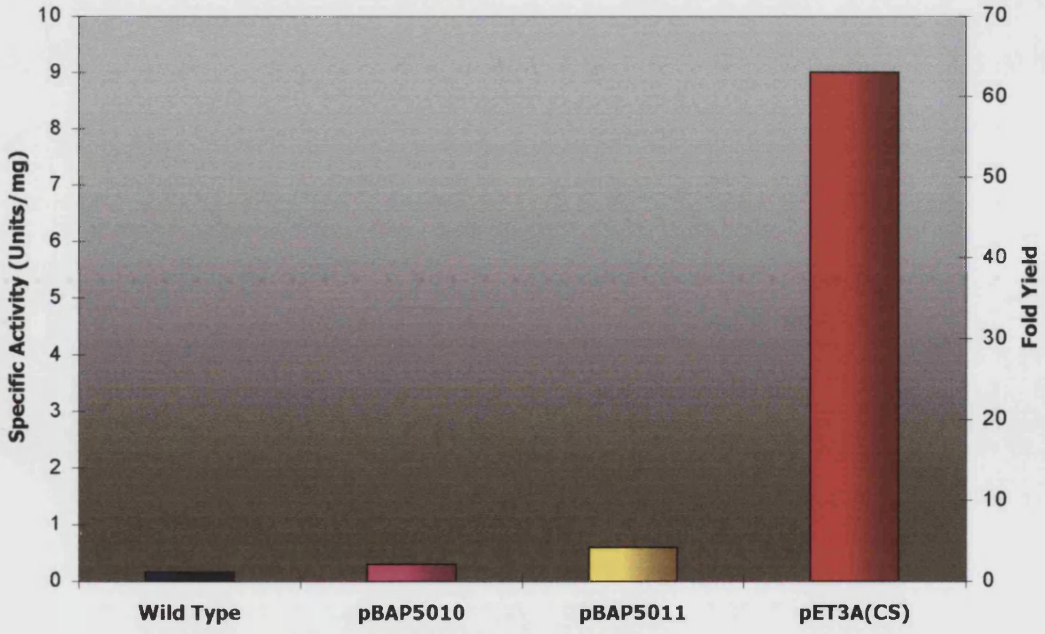
The problems encountered in the construction of the recombinant shuttle vectors pBAP5010 and pBAP5011 stem from not having a copy of the plasmid vector without an insert. Confimation that the DHLipDH gene had been completely removed was made difficult by the fact that the DHLipDH gene, which was being removed, was very similar in M_r to the citrate synthase gene replacing it. Screening for recombinant colonies using PCR enables a large number of colonies to be screened both efficiently and quickly. However, caution must be taken to ensure that both the plasmid and the insert are correct. As was seen in the construction of pBAP5010, most of the recombinant colonies contained the gene of interest which was the parameter being screened for; however, the majority also contained the original cloning vector in which the citrate synthase gene was first cloned.

The construction of pBAP5011 was made easier since the potential pitfalls had already been encountered in the construction of pBAP5010 and PCR screening mechanisms were already in place.

6.4.2 Choosing an Expression System to Produce Recombinant Hf volcanii Citrate Synthase

Active forms of citrate synthase were obtained when the enzyme was expressed both in its native host *Haloferax* and *E.coli*. There were two objectives in choosing an expression system, firstly a system was sought which would produce enough protein to enable crystallisation trials to be set up; secondly, the chosen system was needed to act as tool in the construction and characterisation of mutant enzymes produced by site-directed mutagenesis. The criteria for the first mentioned objective have been met by the pET3A expression system. This is clearly illustrated in Figure 6.25 which shows that the amount of recombinant protein produced from the particular culture volume used is an adequate amount with which to set up crystallisation trials. The amount of protein produced by the pET3A expression system is depicted in comparison with that produced from identical culture volumes from the homologous pBAP expression system.

Figure 6.25 A Comparison of the Levels of Expression of Citrate Synthase from the Haloferax and the E.coli Expression Systems.



Specific Activities refer to the activity of the cell extracts.

The second criterion for selection, an expression system in which to express site-directed mutants has been met in essence by both the homologous and heterologous expression systems; although enzymic activity from protein expressed in the homologous system would always have to be monitored with respect to a non-transformed host cell citrate synthase activity since we do not have a citrate synthase minus strain *of* the host. When we compare various parameters of the proteins expressed and the dynamics of both expression systems , pET3A once again becomes the system of choice. A summary of the salient features of both expression systems is seen in Table 6.5.

Table 6.5 Comparison of Expression of Citrate Synthase in Haloferax and E.coli

Host Organism	***Haloferax* volcanii**		***E.coli****	
Specific Activity (CFE) (Units/mg)	Wild Type	0.15	pET3A (citrate synthase)	9
	PBAP5010	0.3		
	PBAP5011	0.6		
Specific Activity (Pure) (Units/mg)	Wild Type	31	pET3A (citrate synthase)	40
	PBAP5010	62		
	PBAP5011	142		
Promotor on plasmid	PBAP5010	Constitutive rRNA operon *H.cutirubrium*	pET3A (citrate synthase)	Inducible IPTG (0.4mM) Bacteriophage T7
	PBAP5011	rRNA operon + genes own promotor		
Growth Time	PBAP5010 + pBAP5011		pET3A(citrate synthase)	
Appearance of colonies after tranformation	10 days		16 h	
OD_{600} (1-2) in 100ml liquid media inoculated from glycerol stock.	7 days		24 h	
Total growth time	17 days		40 h	
Enzyme reactivation time	0 h		12 h	

*All values quoted for expression in *E.coli* are taken from Connaris *et al.* (1999).

As before, the amount of protein produced was a selection criterion which went in favour of the *E.coli* pET3A expression system. Another selection factor considered was the growth rate of the host organism; even when taking into account the time needed to reactivate the enzyme in the case of the *E.coli* expression system, protein was still expressed approximately ten times faster than in the native system. This was due to the slow growth rate of *Haloferax* cells.

Purification of a halophilic protein in a non-halophilic host is made simpler by the tendency of the host organism proteins to precipitate upon contact with a high salt environment; not only does this assist with purification but also removes host *E.coli* citrate synthase background activity from subsequent enzyme assays. Since we do not have a citrate synthase minus strain of *Haloferax*, background citrate synthase activity would cause problems when characterising citrate synthase site-directed mutants.

6.5 Conclusions

Halophilic enzymes are not always actively expressed in non-halophilic hosts; in the case of the pET3A expression system using *E.coli* as the host organism, expression of CS was achieved. Although the enzyme was not instantly active, full activity was regained by reactivating the enzyme in the presence of 2 M KCl. This expression system was chosen over the native-host pBAP system with the *Haloferax* host because of the rapid growth rate of the *E.coli* host coupled with the increased amount of protein produced.

Chapter 7

Site Directed Mutagenesis of Haloferax volcanii Citrate Synthase

7.1 Introduction

The model of *Hf volcanii* CS as described in chapter 5 highlighted certain physical features that are commonly associated with known structures of halophilic proteins, the most obvious of which is a high surface negative electrostatic potential. The optimisation of the pET expression system (chapter 6), in conjunction with the homology model, have provided the foundations for an investigation into the structural basis of halophilicity by means of site directed mutagenesis.

Comparisons of the halophilic elongation factor (EF-Tu) with its *E.coli* counterpart (Baldacci *et al.*, 1990) have suggested that the binding of hydrated salt ions used to shield the proteins excess negative charge may be via a coordinated mechanism. The co-ordinated binding of hydrated salt ions could be a possible topic of study, but investigation into this hypothesis by means of site-directed mutagenesis would be a monumental tas and one that was not within the time constraints of the project. Another further investigation into the structural basis of halophilicity of the enzyme DHLipDH from *Hf volcanii*, involved the mutation of a single site thought to be a structurally important K^+ binding site found at the intermolecular subunit interface. The stability of the molecule was altered by substituting the chosen residue for various other amino acids (Jolley *et al.*, 1996).

Halophilic proteins need to maintain the spatial alignment of the key residues in the catalytic and binding regions of the core, while under increasing saline conditions the molecule is under pressure to contract from the increased surface tension at the protein-solvent interface. Occupying of such a K^+ binding site could increase the stability of the core, as salt concentration increases.

Following the studies on *Hf volcanii* DHLipDH, a search was made using the *Hf volcanii* CS homology model to look for putative K^+ binding sites that could play an important role in conferring stability of the molecule.

7.2 Design of Mutants

Using the program GRID (Molecular discovery Ltd.), a search was made of the model of *Hf volcanii* citrate synthase to identify potential potassium binding sites. GRID determines energetically-favourable binding sites for a small probe on a target molecule of three-dimensional structure (Goodford, 1985). The strongest potential site was found at the dimer interface of the halophilic CS (Fig 7.1). The site consisted of 2 co-ordinated glutamate (residue 86) and 2 co-ordinated aspartate residues (residue 82), one residue from each respective monomer (Fig 7.2).

In order to select which amino acid was to be mutated, an alignment of citrate synthase sequences of the organisms from the same branch of the phylogenetic tree (Chapter 4) was complied (Fig 7.3). In some of the other sequences included in the alignment, the residue equivalent to position 82 in *Haloferax* CS was also negatively charged; however, in no other case was the negative charge conserved for the equivalent of *Haloferax* residue 86.

Figure 7.1 Position of the Putative K+ Binding Site within the Citrate Synthase Model Structure

The image was produced using the software programme 'O'.
The large domain of each monomer is shown in green and turquoise to highlight the dimer interface. Purple indicates the small domain of each monomer and the putative K+ binding site is depicted in the ball and stick representation.

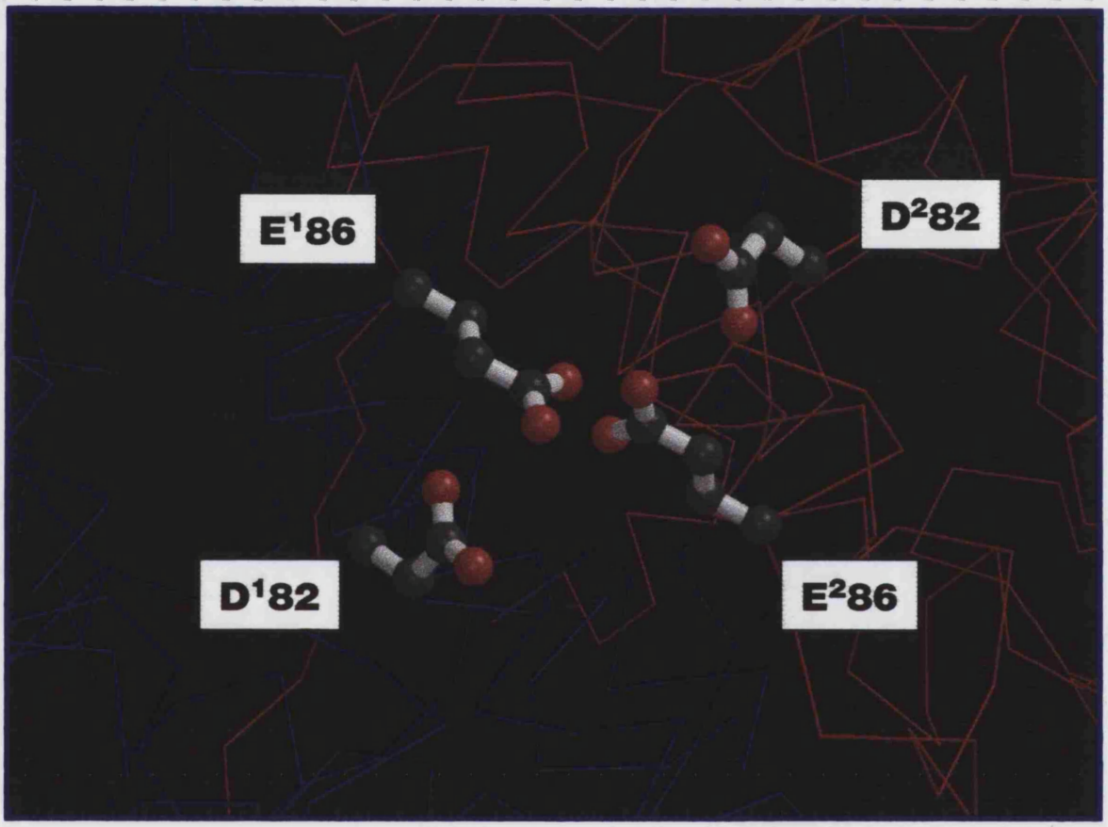


Figure 7.2 The Potential Putative K+ Binding Site.

The site consists of four co-ordinated residues, a glutamate and aspartate from each monomer. Residue E86 was chosen as the best candidate for site directed mutagenesis. The 2 monomeric units are shown in blue and red respectively.

Figure 7.3 Sequence Alignment of the Region to be Mutated

Alignment compiled using PILEUP (GCG) .
Pileup visualised using Genedoc31.
Residues highlighted in red indicate 80% sequence identity.
Residues highlighted in blue indicate 60% sequence identity.

It was therefore decided to mutate residue E86 in order to investigate its possible role in conferring halophilicity. Four mutations were chosen to investigate the effects of residue size and charge on the putative K^+ binding site. The nature of the substitutions is similar to those made in the investigation of the K^+-binding site of DHLipDH; this study was also carried out in our research group (Jolley *et al.*, 1996).

E86A – Alanine is a small aliphatic residue.

E86D – Aspartate conserves the negative charge, but is smaller than Glutamate.

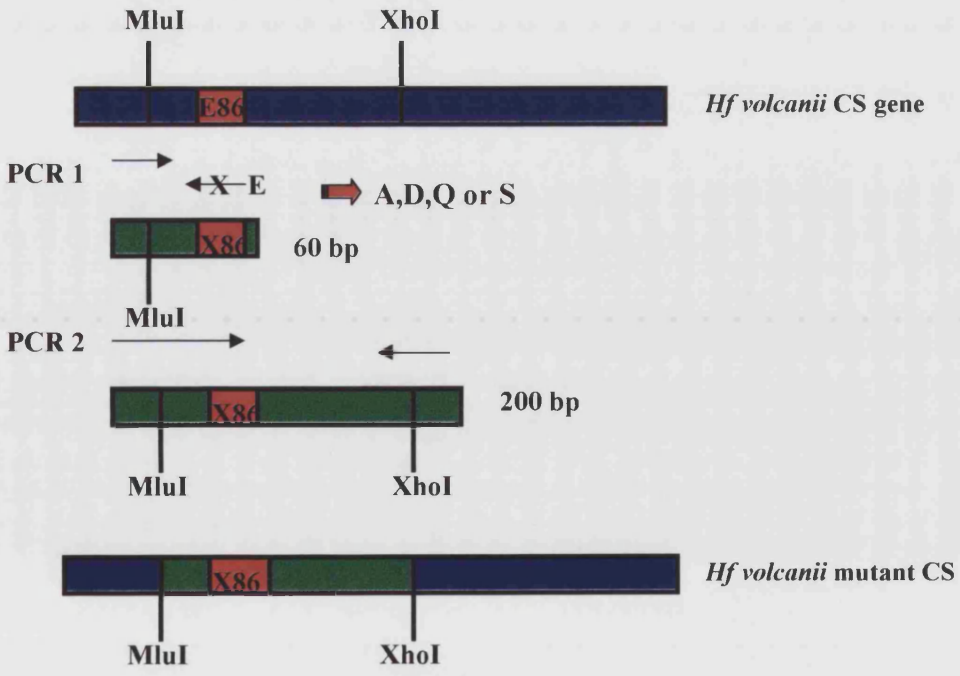
E86Q – Glutamine is similarly sized to glutamate but has no negative charge.

E86S – Serine is smaller than glutamate and is polar.

7.3 Mutagenesis.

Mutagenesis was achieved using a PCR approach making use of the unique restriction sites *MluI* and *XhoI* found flanking the region to be mutated. An initial PCR reaction took place that produced a small product that incorporated both the *MluI* site and the desired mutation. Using the product from this initial reaction as a primer, and an antisense primer that incorporated the *XhoI* site, a second round of PCR was conducted. The PCR fragment of the gene incorporating the desired mutation was restricted with *MluI* and *XhoI* and was then used to replace the wild type fragment simultaneously cut out of the wild type expression vector.

Figure 7.4 Mutagenesis Strategy



7.3.1 Preparation of DNA for mutagenesis

Wild type CS in the pET3A expression system was transformed into *E.coli* (JM109) and the cells were plated out onto LB/agar supplemented with carbenicillin (50 μg/ml). Transformants were selected and used to inoculate a 10ml LB culture which was grown overnight; this in turn was used to sub-culture a larger 100ml LB culture supplemented with carbenicillin from which a plasmid maxiprep (Qiagen) was performed. This pDNA was to act both as a DNA source for mutagenic PCR reactions and to provide the framework of the construct into which the mutagenic regions would eventually be ligated.

7.3.2 Mutagenesis reactions

Four mutagenic primers were designed to change residue E86 from glutamate to the chosen substitute amino acid. Two other non-mutated primers were designed to incorporate the unique restriction sites *MluI* and *XhoI* to be used in the cloning procedure.

Figure 7.5 Primers to Incorporate Mutant sites

E86A	**5'** TGT TCG GCG AGC **GCG** CGT GCC ACG T **3'**
E86D	**5'** TGT TCG GCG AGA **TCG** CGT GCC ACG T **3'**
E86Q	**5'** TGT TCG GCG AGC **TGG** CGT GCC ACG T **3'**
E86S	**5'** TGT TCG GCG AGG **CTG** CGT GCC ACG T **3'**

Nucleotides that have been altered to incorporate restriction site are shown in bold

Figure 7.6 Primers to Incorporate Restriction Sites.

SDMMLU **5'** AGG AAC TCG **A/CG CGT** TCT CCG A **3'**

SDMXHO **5'** TTG CCT T **CT CGA/ G** GT TGA CCT C **3'**

Restriction sites are shown in bold. /Indicates the cut site

An initial PCR reaction was performed using the appropriate mutant primer as an antisense oligomer and SDMMLU as the opposing sense oligomer. Each mutant was amplified in duplicate and in each case a product of the expected size (60 bp) was observed (Fig 7.7). Each PCR product was expected to have incorporated the appropriate mutation and to contain the *MluI* site.

10 μl of each PCR product 1 was used as a sense primer in a second round of PCR using SDMXHO as the antisensense oligomer. PCR was conducted using identical conditions to the initial PCR reactions. Again products of the expected sizes (200bp) were observed in each case; these were electrophoresed on 1% agarose gels and run in unison with the initial PCR product to observe an obvious difference in size (Fig 7.8). These products were expected to contain the unique restriction sites *MluI* and *XhoI* and the relevant engineered SDM mutations.

Samples of relatively low M_r as observed here could normally be better visualised on acrylamide gels; however, they can still be visualised on agarose and were electrophoresed in this manner in order that they might be purified by the Geneclean system for future ligation into the pET vector.

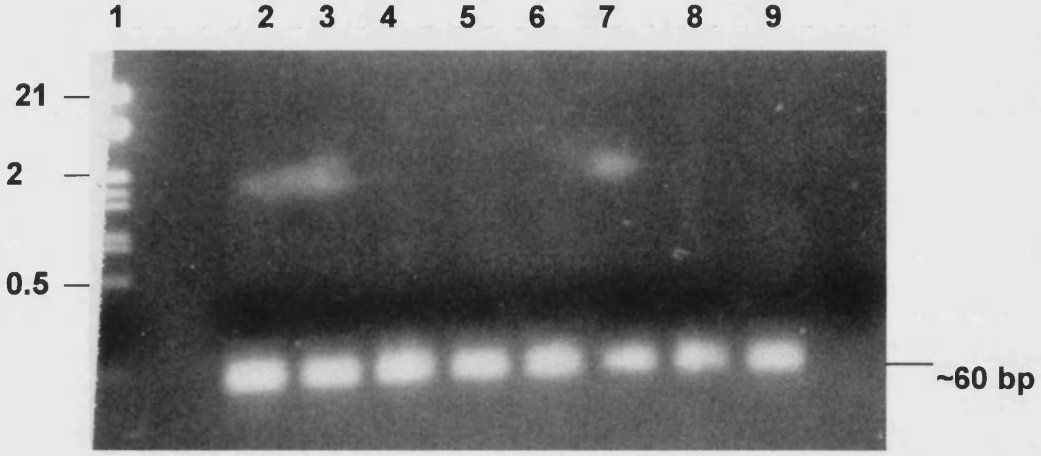


Figure 7.7
Electrophoresis of the Products from the PCR Reaction Used to Incorporate Mutant Sites into the Hf. volcanii Citrate Synthase Gene.

Lane 1: λ/H/E markers
Lane 2: } SDM A
Lane 3: }
Lane 4: } SDM D
Lane 5: }
Lane 6: } SDM Q
Lane 7: }
Lane 8: } SDM S
Lane 9: }

Fragment sizes are indicated in kilobases unless stated otherwise.
The PCR products were electrophoresed on 1% (w/v) agarose gels.

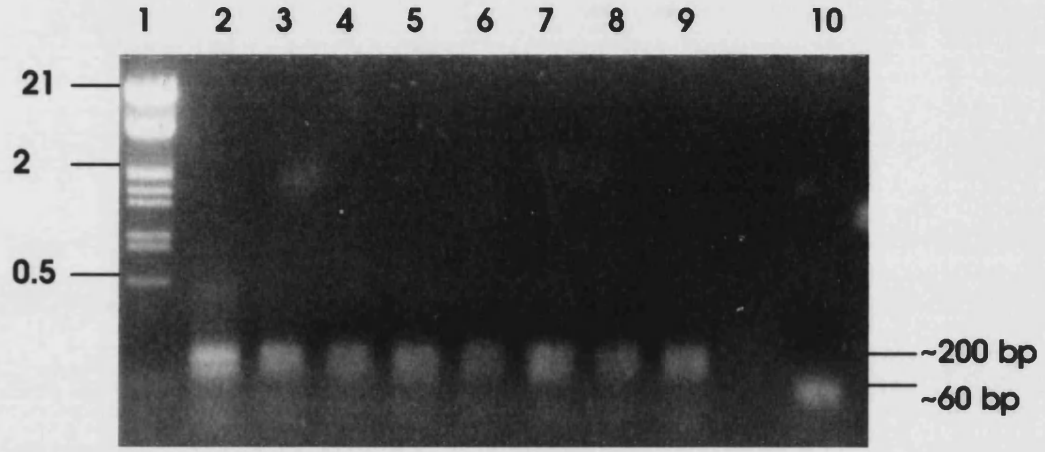


Figure 7.8
Electrophoresis of the Products from the Second PCR Reaction Used to Incorporate Mutant Sites into the Hf. volcanii Citrate Synthase Gene.

Lane 1: λ/H/E markers
Lane 2: } SDM A
Lane 3: }
Lane 4: } SDM D
Lane 5: }
Lane 6: } SDM Q
Lane 7: }
Lane 8: } SDM S
Lane 9: }
Lane 10: 60 bp product from PCR 1 (Fig 7.1)

Fragment sizes are indicated in kilobases unless stated otherwise.
The PCR products were electrophoresed on 1% (w/v) agarose gels.

7.3.3 Selection of Mutants

Incorporation of a different amino acid residue in position 86 slightly alters the restriction map of the enzyme (Table 7.1); this can be utilised for checking the mutation is still present, thereby ensuring that the PCR product has not reverted back to the wild type.

Table 7.1 Restriction Sites that have Altered Within the Gene in the Construction of the Mutant Enzymes

Restriction sites in the wild type enzyme found between the mluI and XhoI sites.			
AciI, BccI, Cac8I, Fnu4HI, HinFI, PleI, TaiI, TaqI, TauI, ThaI			
Restriction sites gained (+) or lost (-) in SDM primers			
E86A	**GCG**	*+BssHII, HhaI*	
E86D	**GAT**	*+DpnI, Sau3AI*	
E86Q	**CAG**	*+AceIII, AluI, CviJI,*	*-ThaI*
E86S	**AGC**	*+BbvI, CviJI, MnlI, TseI*	*-ThaI*

Products from the second PCR reaction were restriction digested to ensure that the correct mutations had been incorporated. E86A, D, and S were restricted with *BssHII*, *Sau3AI* and *BbvI* respectively (Fig 7.9), as additional sites for these enzymes have been incorporated through the SDM. E86Q was restricted with *ThaI* as the recognition sequence for this enzyme position 86 was destroyed as a consequence of the mutation. However, *ThaI* was still expected to restrict the 200bp fragment, producing a larger digested fragment than the wild type, due to another site still being present in the PCR product (fig 7.9).

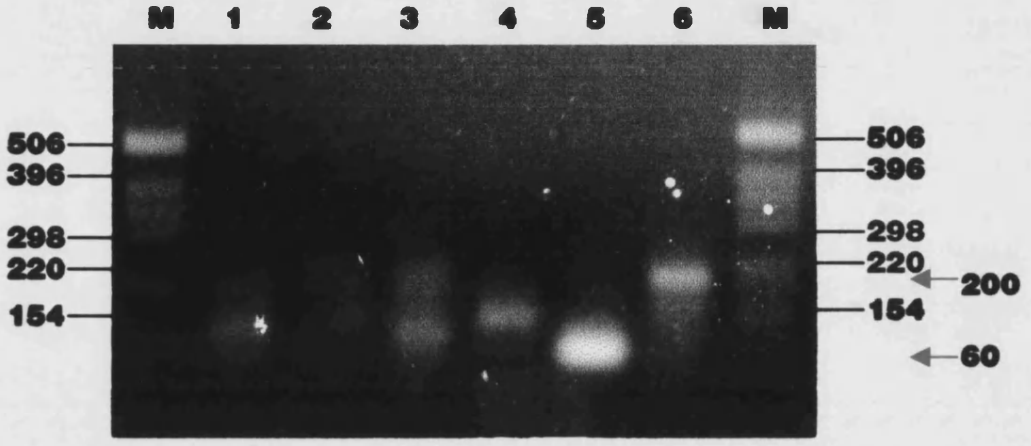


Figure 7.9
Digests performed on stage 2 (200bp) PCR fragments (Fig 7.8) to ensure correct mutations have been incorporated.

Lane M: λ/H/E markers
Lane 1: E86 A PCR product digested with BssHII
Lane 2: E 86 D PCR product digested with Sau3AI
Lane 3: E 86 S PCR product digested with BbVI
Lane 4: E 86 Q PCR product digested with ThaI
Lane 5: 60 bp PCR product from stage 1 PCR
Lane 6: 200 bp PCR product from stage 2 PCR prior to digestion

Fragment sizes are indicated in basepairs unless stated otherwise.
The PCR products were electrophoresed on a 1% (w/v) agarose gels

7.3.4 Ligation of the Mutant PCR Fragment into the pET Vector

Restriction digests using *MluI* and *XhoI* were performed on the pET3A vector containing the wild type CS gene to remove the region to be replaced by the mutated PCR fragment (Fig7.10). Single digests were also performed to ensure that the individual enzymes were still active. As the excised region was of a small M_r, and not visible after electrophoresis (Fig 7.10), another digest using *MluI* and *BamHI* was performed in unison to cut out a fragment that was visible on an agarose gel to check that the plasmid still contained the CS gene.

The *MluI/XhoI* double digested pET vector containing the CS gene was removed from the agarose gel and purified using Geneclean™. Ligations using this linearised vector and each of the purified 200bp PCR products were set up and the resulting products were transformed into *E.coli* (JM109) and the cells plated onto LB/agar containing carbenicillin.

Potential transformants were selected and used to inoculate LB media. SNAP™ plasmid minipreps were used to produce DNA, which was once again restricted to check for the presence of the mutant amino acid residue. The mutants E86A,D,Q and S were restricted with *BssHII, Sau3AI, AluI* and *BbvI* respectively (Fig 7.11). E86A was the only mutant that incorporated a restriction site unique to both the CS gene and the pET3A vector. Single digests were performed on the complete plasmid. The plasmid containing the unmutated CS gene provided a control in each case.

In the case of the mutant E86A it is obvious that the plasmid contains the mutation due to the presence of the unique restriction site. However, in the other three mutants the results of the digestion were not as clear; therefore the presence of each mutation was confirmed by sequencing. The plasmids were sequenced using the primer (Hcforward; Table 2.11), which was initially used in the original construction of the pET3ACS clone (Connaris *et al.*, 1999).

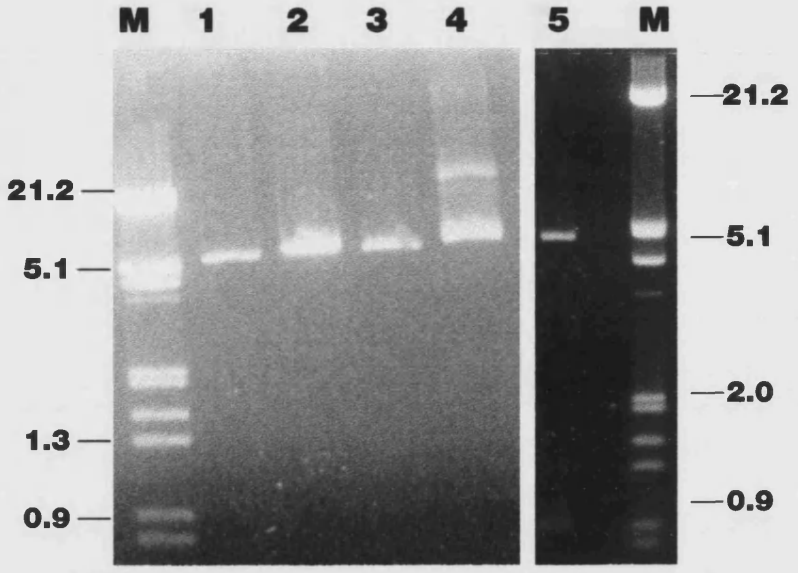


Figure 7.10
Digests performed on pET3A shuttle vector containing CS construct.

Lane M: λ/H/E markers
Lane 1: pET3A digested with MluI & XhoI
Lane 2: pET3A digested with MluI
Lane 3: pET3A digested with XhoI
Lane 4: undigested pET3A
Lane 5: pET3A digested with MluI & BamHI

The DNA was electrophoresed on 1% (w/v) agarose gels.
Fragment sizes are indicated in kilobases.
pET3A refered to above is the commercial pET3A vector containing the CS construct.

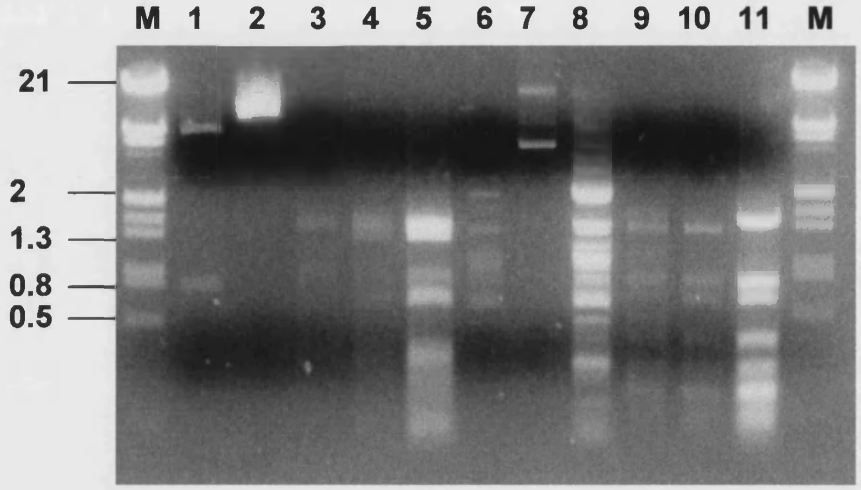


Figure 7.11
Gel of Digestion of pET3A Vectors thought to Contain CS Gene with Engineered SDM Mutations.

Lane M: λ/H/E markers
Lane 1: SDM A
Lane 2: WT pET3A(CS) } **Digested with BssHI**
Lane 3: SDM D1
Lane 4: SDM D2
Lane 5: WT pET3A(CS) } **Digested with Sau3AI**

Digested with AluI { **Lane 6: SDM Q1**, **Lane 7: SDM Q2**, **Lane 8: WT pET3A(CS)**
Digested with BbvI { **Lane 9: SDM S1**, **Lane 10: SDM S2**, **Lane 11: WT pET3A(CS)**

Fragment sizes are indicated in kilobases .
The PCR products were electrophoresed on 1% (w/v) agarose gels.

The complete fragment that was generated by PCR was confirmed by sequencing to ensure that the *MluI* and *XhoI* restricted regions had not changed and that no PCR induced errors had been incorporated. The specific mutated regions incorporating the required nucleotide changes can be seen in Figure 7.12.

7.4 Heterologous Expression of the Mutant Enzymes

The resulting plasmids containing the correct mutations were transformed into *E.coli* BL21 (DE3) cells for expression. Over-expression of the citrate synthase gene was obtained with cells grown at 30°C in 100ml LB containing carbenicillin (50μg/ml). Induction was achieved at 25°C by adding IPTG to a final concentration of 0.4 mM at OD_{600} 0.6. Cultures were left shaking overnight before the cells were harvested by centrifugation.

Cell pellets from the mutant cultures were resuspended in 20 mM Tris-HCl buffer, pH8 , containing 1 mM EDTA and 2 M KCl before sonication and centrifugation at 10,000g for 10 min. Due to the heterologously-expressed enzyme needing time to reactivate in the presence of 2 M KCl, the cell extract was kept at 4°C overnight.

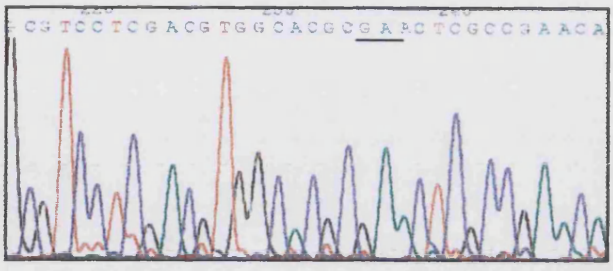
7.5 Purification of the Mutant Enzymes

It was not taken for granted that the mutant and wild type enzymes could all necessarily be purified in the same manner. However, preliminary small-scale pilot investigations showed that all mutants did in fact bind to Matrex Gel Red A in a similar fashion to the wild type enzyme.

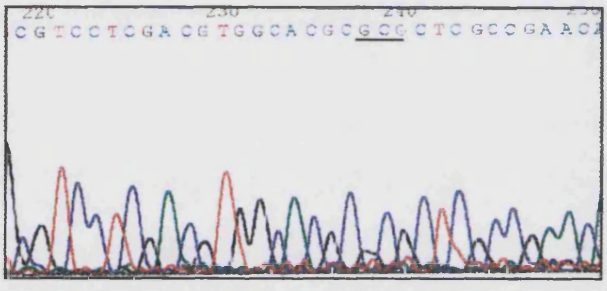
The activated cell extract was diluted 10 fold in 10 mM Tris-HCl, pH 8 , containing 0.1 mM EDTA prior to purification. The citrate synthases were purified using the method outlined in chapter 6, and the resulting enzyme solutions were brought to 2 M salt by the addition of KCl and were stored at 4°C. Fractions were collected and assayed for citrate synthase activity. Figures 7.13 – 7.16 shows SDS-PAGE gels of the purification step; yields and specific activities from the procedure can be seen in Table 7.2.

Figure 7.12
Conformation of Incorporation of New Amino Acid in Position 86 by Automated Sequencing.

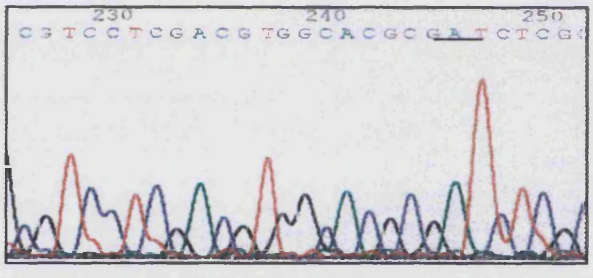
WT
Glutamate-E -GAA



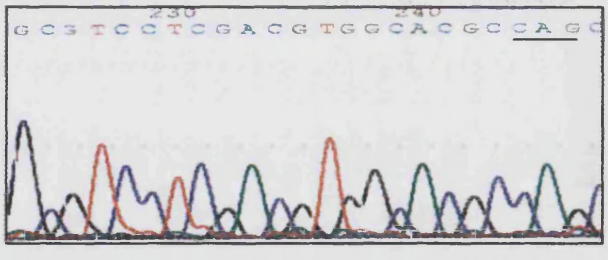
SDM A
Glutamate ➡Alanine
E ➡A GAA➡ GCG



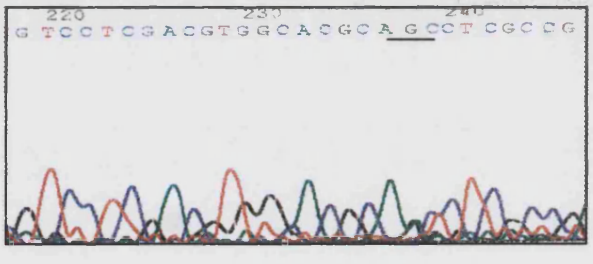
SDM D
Glutamate ➡Aspartate
E ➡D GAA➡ GAT



SDM Q
Glutamate ➡Glutamine
E ➡Q GAA➡ CAG



SDM S
Glutamate ➡serine
E ➡S GAA➡ AGC



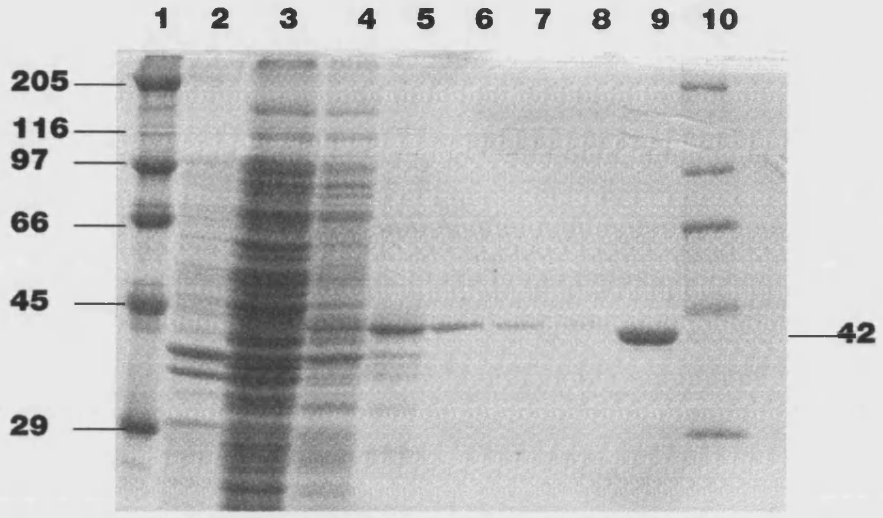


Figure 7.13
SDS-PAGE Analysis of the purification from E.Coli of the recombinant mutant CS (E86A)

Lane 1 &10: Broad range molecular weight markers
Lane 2: Insoluble cell extract from untransformed host
Lane 3: Soluble cell free extract from untransformed host
Lane 4: Insoluble cell extract from transformed host
Lane 5: Soluble cell free extract from transformed host
Lanes 6-8: Fractions collected after Red Gel A purification
Lane 9: *Hf. volcanii* CS standard from previous purification donated by Dr H. Connaris

Protein sizes are indicated in kilodaltons.
Proteins were electrophoresed on a 10% SDS polyacrlyamide gel.

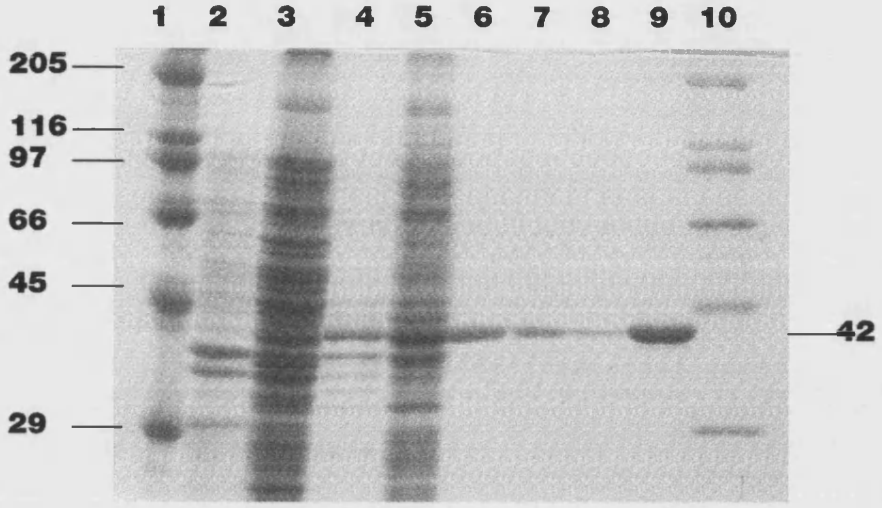


Figure 7.14
SDS-PAGE Analysis of the purification from E.Coli of the recombinant mutant CS (E86D)

Lane 1 &10: Broad range molecular weight markers
Lane 2: Insoluble cell extract from untransformed host
Lane 3: Soluble cell free extract from untransformed host
Lane 4: Insoluble cell extract from transformed host
Lane 5: Soluble cell free extract from transformed host
Lanes 6-8: Fractions collected after Red Gel A purification
Lane 9: *Hf. volcanii* CS standard from previous purification donated by Dr H. Connaris

Protein sizes are indicated in kilodaltons.
Proteins were electrophoresed on a 10% SDS polyacrlyamide gel.

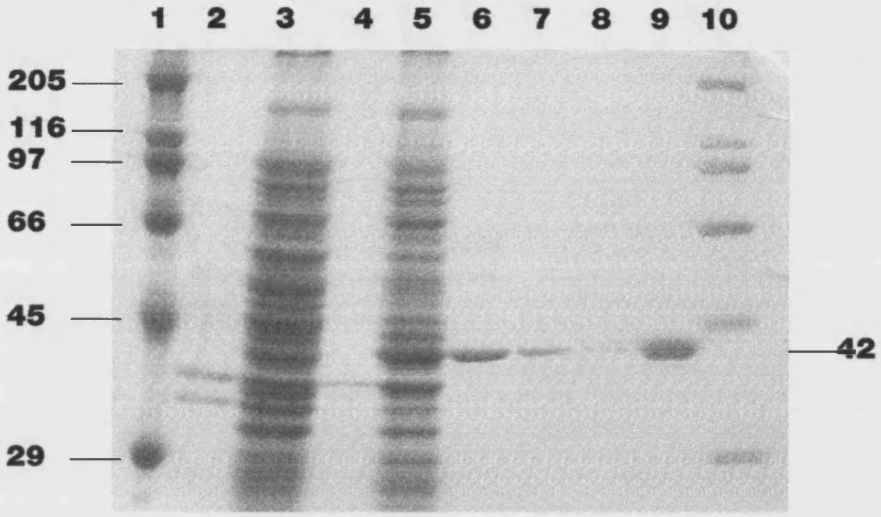


Figure 7.15
SDS-PAGE Analysis of the purification from E.Coli of the recombinant mutant CS (E86Q)

Lane 1 &10: Broad range molecular weight markers
Lane 2: Insoluble cell extract from untransformed host
Lane 3: Soluble cell free extract from untransformed host
Lane 4: Insoluble cell extract from transformed host
Lane 5: Soluble cell free extract from transformed host
Lanes 6-8: Fractions collected after Red Gel A purification
Lane 9: *Hf. volcanii* CS standard from previous purification donated by Dr H. Connaris

Protein sizes are indicated in kilodaltons.
Proteins were electrophoresed on a 10% SDS polyacrlyamide gel.

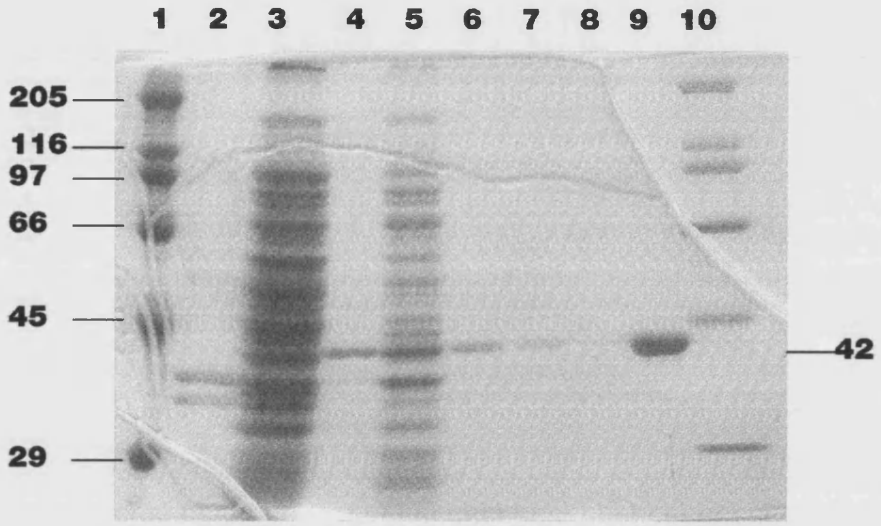


Figure 7.16
SDS-PAGE Analysis of the purification from E.Coli of the recombinant mutant CS (E86S)

Lane 1 &10: Broad range molecular weight markers
Lane 2: Insoluble cell extract from untransformed host
Lane 3: Soluble cell free extract from untransformed host
Lane 4: Insoluble cell extract from transformed host
Lane 5: Soluble cell free extract from transformed host
Lanes 6-8: Fractions collected after Red Gel A purification
Lane 9: *Hf. volcanii* CS standard from previous purification donated by Dr H. Connaris

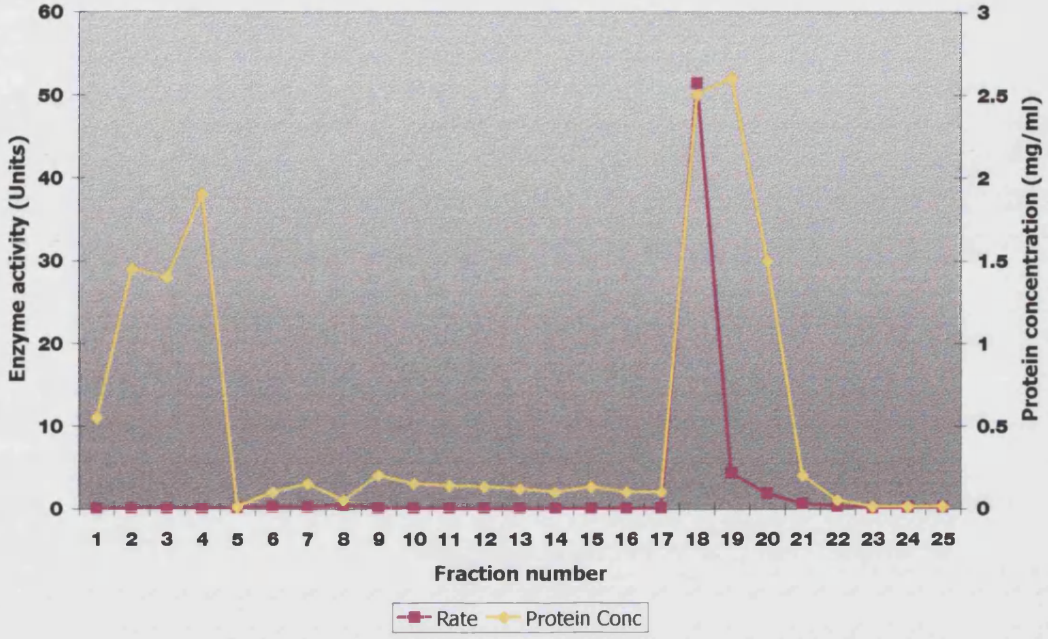
Protein sizes are indicated in kilodaltons.
Proteins were electrophoresed on a 10% SDS polyacrlyamide gel.

Table 7.2 Purification Table for Mutant and Recombinant Wild-Type Halophilic Citrate Synthase.

Purification Step	Total [Protein] (mg)	Total Activity Units	Specific Activity Units/mg	Yield (%)
Wild-type				
CFE	28	240	8.6	100
Matrex Gel Red A	3	247	82	103
SDM A				
CFE	54	360	6.7	100
Matrex Gel Red A	4.8	291	60	81
SDM D				
CFE	41	120	3	100
Matrex Gel Red A	1.6	115	72	96
SDM Q				
CFE	41	200	5	100
Matrex Gel Red A	5.6	320	57	160
SDM S				
CFE	46	160	3.4	100
Matrex Gel Red A	2.6	227	87	142

Samples were assayed in 10 mM Tris (pH 8), 1 mM EDTA containing 2 M KCl at 30 °C

Figure 7.17 Elution Profile of the Mutant E86A from Matrex Gel Red A



From the elution profile of E86A from the Matrex Gel column, it can clearly be seen that approximately 50% of the cell protein was washed out in fractions 1-4, and that elution of the enzyme occurred in fraction 18 immediately after the addition of elution buffer containing of 0.2 mM CoA, 1mM OAA. Unbound citrate synthase activity was minimal, implying that the majority of the enzyme bound to the column. Almost all of the enzyme was removed from the column with the addition of the elution buffer, reflecting the high recovery yields obtained (Table 7.2). The elution profile was found to be representative of the other mutant and the wild-type enzymes.

7.6 Characterisation of mutant Enzymes

7.6.1 Kinetic parameters

Kinetic parameters were determined in 10 mM Tris (pH8), 1 mM EDTA containing 2 M KCl at 30 °C. (Table 7.3) The direct linear plot of Eisenthal and Cornish-Bowden (1974) was used to determine K_m and V_{max}, values and the data were processed using the programme Enzpack 3

Table 7.2 Kinetic Parameters of Recombinant Wild-Type and Mutant Halophilic Citrate Synthase.

	K_m (OAA) (μM) (±SE)	K_m (ACoA) (μM) (±SE)	V_{max} (U/mg)
Wild Type	54 ± 3	107 ± 8	72 ± 2
E86A	53 ± 1	90 ± 5	90 ± 2
E86D	40 ± 3	119 ± 26	116 ± 3
E86Q	48 ± 2	122 ± 6	94 ± 3
E86S	44 ±3	148 ± 13	87 ± 2

All mutants are slightly more active than the recombinant wild-type enzyme; the increase is small in most cases with the exception of E86D which has a 60% increase in catalytic activity. K_m values for both substrates show no significant deviation from the wild-type values, indicating that the mutant enzyme affinity for both substrates remains unaltered.

7.6.2 Enzyme Efficiency

Table 7.4 K_{cat}/K_m for Both Citrate Synthase Substrates

	K_{cat} (min^{-1})	K_{cat}/km OAA (M^{-1} min^{-1})	K_{cat}/km AcoA (M^{-1} min^{-1})
WT	2800	1×10^{8}	3×10^{7}
E86A	3780	7×10^{7}	4×10^{7}
E86D	4872	1×10^{8}	4×10^{7}
E86Q	3948	6×10^{7}	3×1^{7}
E86S	3654	8×1^{7}	3×10^{7}

7.6.3 Salt Activity Profiles

Salt activity profiles of the mutant enzymes were determined in 10 mM Tris (pH8), 1 mM EDTA containing various concentrations of salt. Assays were performed in both NaCl (Figure 7.18) and KCl (Figure 7.19).

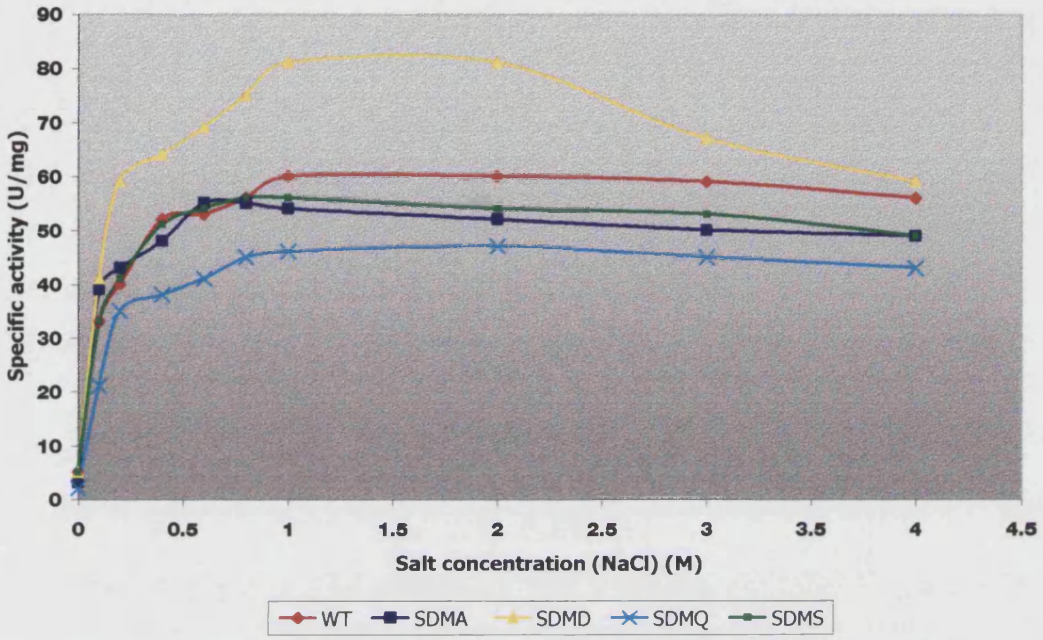


Figure 7.18: Mutant Enzyme Activity Profiles in NaCl

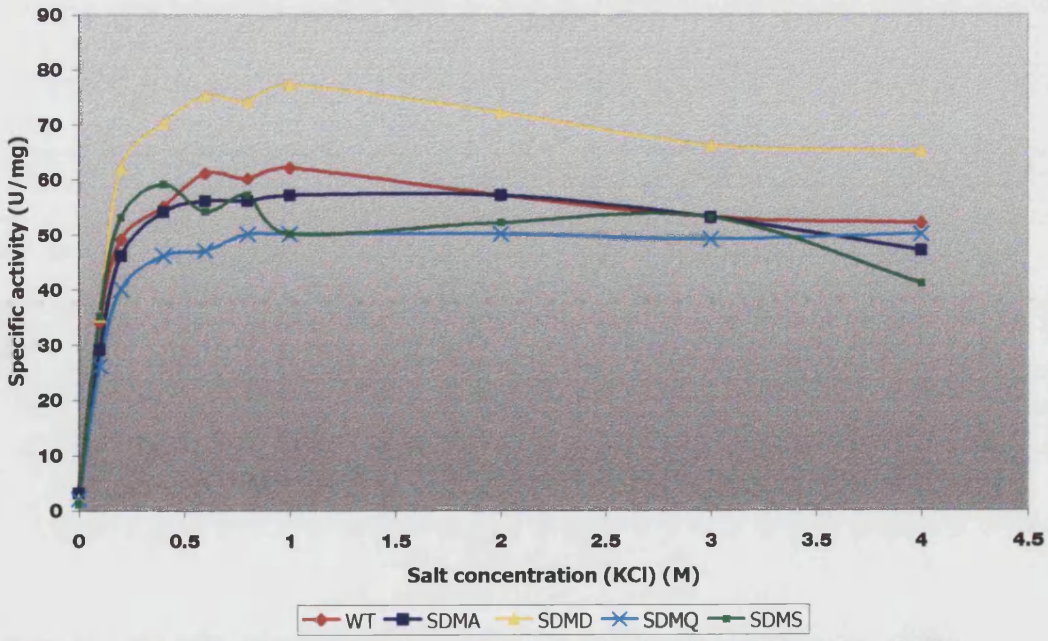


Figure 7.19: Mutant Enzyme Activity Profiles in KCl

7.6.4 Thermal Stability

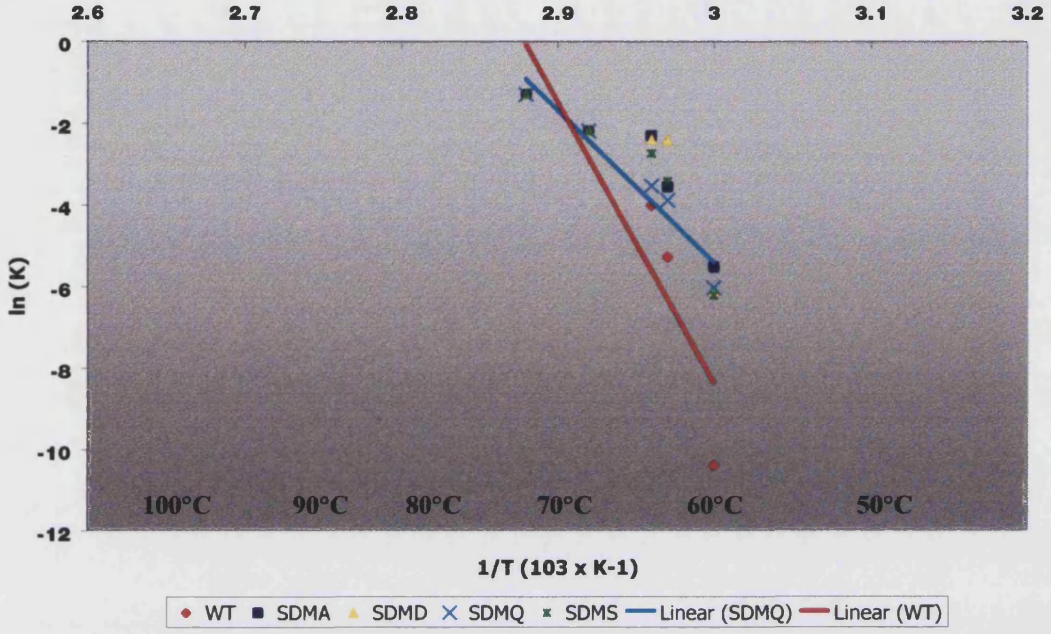
Thermal stability studies were conducted in Epps buffer in the presence of 2M KCl. Lp4 tubes containing 900 μl buffer were heated in a thermostatically-controlled water bath. The temperature in a control tube was monitored constantly, and when it reached the required temperature 100 μl of the undiluted enzyme was added, mixed, and the the timing started. 100 μl alliquots of the enzyme were removed at the alloted time intervals. A t=0 sample was determined by making an identical dilution of enzyme into cold buffer. Thermal inactivation was quenched by rapid cooling on ice. The tubes were briefly centrifuged before assaying at 30°C in Epps containing 2M KCl.

Inactivations were carried out at 60, 63.5, 65, 69 and 74 °C. Activity values were expressed as a percentage of the t=0 value. Logarithmic plots of the inactivations were used to determine the 1st order rate constants (k) of inactivation (Table 7.5), which were used to display the data on an Arrhenius plot.

Table 7.5 1st Order Rate Constants (k) of Inactivation of Mutant and Recombinant Wild-type Halophilic Citrate synthase Enzymes.

	60 °C	63.5 °C	65 °C	69 ° C	74 °C
WT	3×10^{-5}	5×10^{-3}	2×10^{-2}	2×10^{-1}	2×10^{-1}
E86A	4×10^{-3}	3×10^{-2}	10×10^{-2}	2×10^{-2}	2×10^{-1}
E86D	2×10^{-3}	9×10^{-2}	9×10^{-2}	1×10^{-1}	2×10^{-1}
SDMSQ	2×10^{-3}	2×10^{-2}	3×10^{-2}	2×10^{-1}	2×10^{-1}
E86S	2×10^{-3}	3×10^{-2}	6×10^{-2}	2×10^{-1}	2×10^{-1}

Figure 7.20 Arrhenius Plot for the Thermal Inactivation of Wild-type and Mutant Halophilic Citrate Synthase.



For clarity, linear regressions are shown only for wild type and E86Q.

At higher temperatures (69°C and 74°C) the thermal stability properties of all mutants are comparable to the wild-type enzyme. However, at lower temperature (60 – 65 °C) the mutants are all significantly less stable than the wild-type.

7.7 Discussion

7.7.1 Construction of Mutants

The incorporation of the mutant amino acid residues by PCR went quite smoothly with no major problems. In both rounds of PCR required to incorporate the mutation, products of the correct size were gained at the first attempt, they were subsequently shown to be the correct mutations, which was confirmed by sequencing.

6.7.2 Mutant Enzyme Properties.

Caution must be taken when making any structural predictions about the mutant enzymes, as all predictions to date have been based upon a homology model. Preliminary predictions made on the model do not show any gross conformational changes on introduction of the mutations. However, since none of the mutant residues were larger than the wild-type amino acid residue, large structural changes would probably not be expected to show up, although a change in charge could cause a change in conformation.

Kinetic analysis did show an increase in Vmax. However, compensatory changes in Km resulted in similar catalytic efficiencies (kcat/Km).

Salt profiles for both mutants and recombinant wild-type are very similar, with all enzymes achieving full activity in 0.8M salt. The activity profiles remained similar in both KCl and NaCl. The activity profiles (Fig7.16 & 7.17) were not conducted at the time of purification (Table 7.2), but approximately six weeks after. It is clear from the salt profiles that all of the mutants and the wild-type, with the exception of E86D, have lost a significant amount of activity over a time period. The specific activity of E86D remains unaltered; this is not too surprising since a change from glutamate to aspartate was the most conservative mutation

made in terms of electrostatic potential. The conservation of the negative charge in conjunction with a decrease in residue size has appeared to increase the long-term stability of the enzyme.

Thermal stability studies showed that the recombinant wild-type enzyme had a greater thermal stability than the mutant enzymes. Originally, site E86 was chosen to perform site-directed mutagenesis on as it was predicted to be a putative K^+ binding site, and thought to play a role in conferring halostability. As the results have shown, disruption of this site hasn't greatly affected the salt dependency properties of this enzyme.

A similar study involving the site-directed mutagenesis of a putative K^+ binding site in DHLipDH from *Haloferax volcanii* was conducted by Jolley (1996). The results from this study showed that a change from glutamate (E) to either glutamine (Q), alanine (A) or serine (S) in the chosen mutated site significantly reduced enzyme activity. At lower salt concentrations, the native enzyme rapidly lost activity while the mutant enzymes activity remained unchanged; it was inferred that the respective site was involved in conferring halostability, however, this does not seem to be the case with the citrate synthase enzyme.

The slight increase in thermal stability shown by the *Haloferax* wild type enzyme could indicate that residue E86 plays a role in conferring thermal stability of the enzyme rather than an active role in governing catalytic activity. This potential thermal stability could be achieved by involvement in an ionic network. The presence of ionic networks has been regarded as a thermostabilising feature in some thermophilic enzymes (Yip *et al.*, 1998). It has been shown that the ion-pair networks become more fragmented as the temperature stability of the enzyme decreases and are consistent with a role for the involvement of such networks in the adaptation of enzymes to extreme temperatures (Yip *et al.*, 1998). The presence of ionic networks contributing to thermostability have recently been discovered in halophilic proteins (Zaccai personal communication). Analysis of the model of *Haloferax* CS has highlighted the presence of a potential ionic network within each monomer (Figure 7.19). This is comprised of the residues D82 and E86; originally thought to be part of a K^+ binding site, and R85.

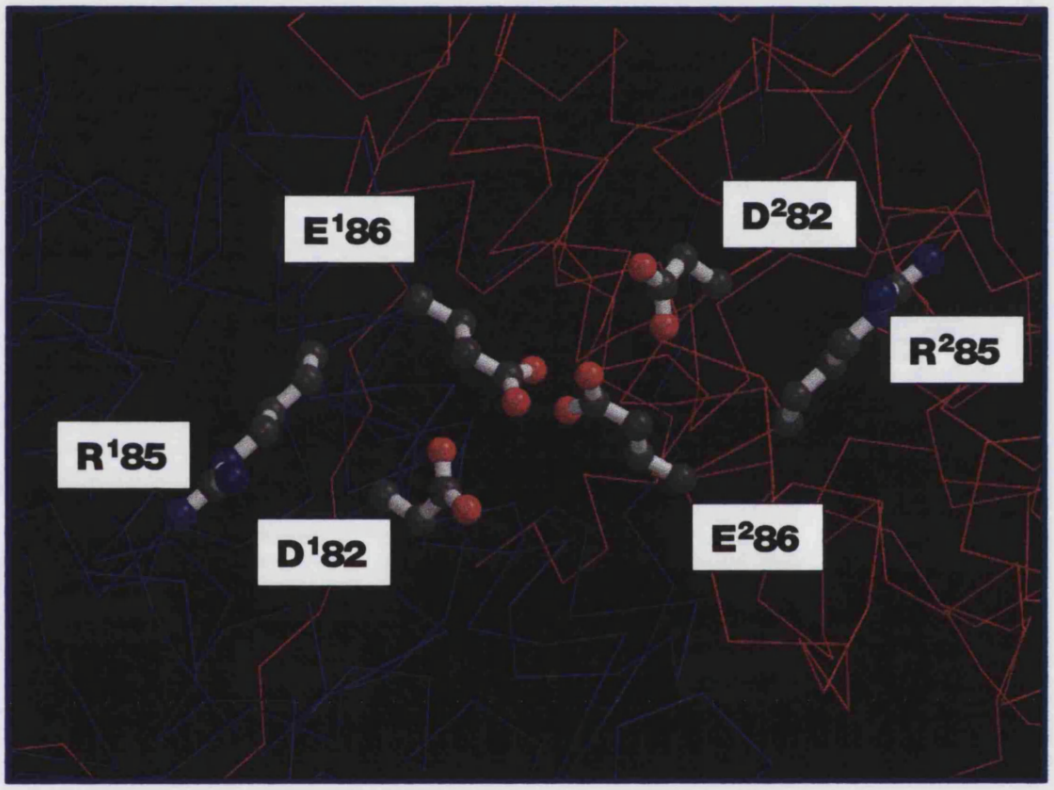


Figure 7.21: The Proposed Ionic Network.

The site consists of six coordinated residues, a glutamate and aspartate as seen previously (Figure 7.2), and an arginine in position 85. One copy of each residue from each respective monomer is thought to contribute to the network.

However, I reiterate that all predictions made are based solely on a homology model in which the spatial positions of the residues in the model may differ from those of the true structural conformation of the protein. This once again highlights the need for a crystal structure to confirm or refute the structural predictions already made on the protein.

Crystallisation trials have been conducted on the wild type enzyme but have been unsuccessful to date. A summary of the crystallisation conditions tried that led to precipitation of the protein are outlined in Appendix 4.

I would like to thank Dr. Rupert Russell for running the GRID analyses and for his assistance in producing the molecular model.

Chapter 8

General Discussion

8 General Discussion

8.1 Expression of Halophilic Proteins

Heterologous expression of foreign proteins is often a compromise between rapid growth, high cell densities and detailed knowledge of the surrogate genetic system versus incorrect post-translational changes, mis-folding of the protein rendering it inactive and the formation of insoluble inclusion bodies. However, heterologous expression of *Haloferax* citrate synthase in this instance was the way forward due to the high levels of active enzyme expressed. Although initially the enzyme was expressed in *E.coli* at low salt concentrations and was therefore inactive, 24 h in the presence of 2M KCl reactivated the enzyme to its full activity. The enzyme was expressed X times more than in the expression system set up in it's native host. The IPTG inducible promoter present on the commercial vector used allowed a level of control over expression of the protein. This enabled the cells to be grown to reasonable cell densities before their translation machinery was exposed to the production of the foreign protein.

Heterologous expression of halophilic proteins has another advantage when it comes to large scale processes. The saline growth media utilised by halophilic organisms are highly corrosive to stainless steel, a commonly used material in large scale fermenters. For large scale production of homologously expressed halophilic proteins specialist fermenters would have to be employed made of a corrosion resistant material.

With our increasing knowledge of halobacterial genetic systems and the broad range of heterologous expression systems available on the market today, heterologous expression of halophilic proteins is becoming increasingly popular. However, there are still difficulties using a system for expressing a protein whose native environment is highly saline. Not all halophilic enzymes expressed in e.coli can be activated by the simple addition of NaCl. In some cases,

refolding from urea is required, in other cases, successful activation has not been achieved (Connaris *et al.*, 1999).

8.2 The Halophilic Hypothesis

It is now generally accepted that halophilic Archaea establish an osmotic balance by maintaining internal levels of salt isotonic with the external high salt environment. The known structures of halophilic proteins all have a high acidic amino acid content the majority of which are found spatially on the protein's surface. The arrangement of these acidic residues is thought to maintain enzyme activity and solubility by attracting water molecules back to the protein's surface. Additional internal hydrophobic interactions are also thought to play an important role in helping these proteins to cope with high salt stress.

The findings of this study confirmed initial observations on the structural basis of protein halophilicity. The *Hf volcanii* citrate synthase sequence has a characteristic high proportion of negatively charged amino acids, which when put into a structural context are predominantly positioned on the protein's surface. The potential potassium binding site identified at the dimer subunit interface of the molecular model proved not to play such a crucial role in conferring the protein's halophilicity and stability as was initially thought. Although this highlights the uncertainties of predictions made on model structures, it is still valid to make such predictions as long as one is aware of the potential positional uncertainties. Until the crystal structure of this enzyme is obtained, it will be difficult to make further progress in undrstanding the molecular basis of halophilicity.

8.3 Future Work on this Project

Future work on this project will concentrate on obtaining protein crystals of suitable quality for X-ray diffraction, and resolving the crystal structure of the enzyme. Numerous crystallisation trials have been conducted but as of yet without success. The various crystallisation conditions already undertaken are outlined in appendix 4.

A crystal structure of a halophilic citrate synthase will complement the suite of crystal structures already obtained from organisms spanning a range of optimum growth temperatures. A confirmed structure would also shed light on the hypothesis already formulated around the molecular model and provide the basis for future structural based studies.

The binding of water to the surface of halophilic proteins has been speculated to be co-operative (Baldacci, *et al.*, 1990). The time constraints on this project did not leave enough time to investigate this phenomenon. A site directed mutagenesis study to look at this hypothesis by creating enzymes with multiple mutations which are structurally placed close to each other on the surface of the protein would be a future possibility. An alternative approach to mutagenesis would be to incorporate random mutagenesis and screen for halophilic mutants by means of directed evolution (Kuchner *et al.*, 1997, Arnold *et al.*, 1996). This approach would probably be a quicker way to pinpoint residues that play an important role in conferring halophilicity.

An additional future task would be to look at the possibility of making chimeric enzymes. This has already been investigated with two thermophilic citrate synthases from *Thermoplasma acidophilum* and *Pyrococcus furiosus,* to try and engineer increased thermostability into an enzyme. A citrate synthase enzyme constructed from *Pyrococcus* and *Haloferax* would be an interesting chimeron to make. One would be able to study the effects on halophilicity and thermostability. The production of an active hybrid enzyme is potentially feasible, for activity in halophilic media the high proportion of negatively charged residues would need to be maintained. Feasibility is also enhanced by the high sequence identity (45%) between the two enzymes.

8.4 Reshaping the Future of Cloning Projects

The publication of the first whole initiated a snowball effect in the sequencing of other genomes and a change to the approach of cloning genes. Since then many other microbial whole genomes have followed including examples from all three domains of life. To date there are six complete archaeal genome sequences available in the databank *Aeropyrum pernix* (Kawarabayasi *et al.*, 1999), *Archaeoglobus fulgidans* (Klenk *et al.*, 1997), *Methanobacterium thermoautotrophicum* (Smith *et al.*, 1997), *Methanococcus jannaschii* (Bult *et al.*, 1996), *Pyrococcus abyssi* and *Pyrococcus horikoshi* (Kawarabayasi *et al.*, 1998) and numerous others in various stages of completion. The sequencing of the *Haloferax volcanii* was one of the first archaeal sequencing projects to get underway, however, no information to date has been released. This may reflect the difficulties in sequencing halophilic genomes especially those with a high G + C content.

The availability of these genetic blueprints is going to phase out cloning and sequencing of specific genes. Future projects such as this one will in essence begin almost where this one left off with a search of a genome database. The future emphasis will focus more on the structural, biochemical and mechanistic rather than the cloning aspect. Naturally this will take time and a genome database can only be searched for genes of interest if the genome has been completely sequenced.

With a growing industrial interest in "novel" uncultured organisms or environmental samples, the organisms are unknown and therefore the old cloning and sequencing routine will probably prevail for a few more years. However with the progression of technology and improvements in the quality and speed of sequencing technology, it probably won't be long before whole genomes are sequenced to order in a short space of time.

References

Altekar, W., Kristjansson, H., Ponnamperuma, C. & Hochstein,L. (1984) On archaebacterial ATPase from *Halobacterium saccharovorum. Ori. Life. Evol. Bios.* **14**: 733-738

Anderson, S.C.K., Powrie, R. & Mitchell C.G. (1994) *Pseudomonas* citrate synthase - purification and characterization. *Biochem. Soc. Trans.* **22:** S41

Arnold, F.H. (1996) Directed evolution: creating biocatalysts for the future *Chem. Eng. Sci.* **51:** 5091-5102

Baldacci, G., Guinet, F. Tillit, J., Zaccai, G. & Recondo, A.M. (1990) Functional implications related to the gene structure of the Elongation factor-EF-TU from *Halobacterium-marismortui. Nucleic Acid. Res.* **18**: 507-511

Barns, S.M., Delwiche, C.F., Palmer, J.D., & Pace, N.R. (1996) Perspctives on archeael diversity thermophily and monophyly from environmental rRNA sequences. *Proc. Natl. Acad. Sci. USA.* **93**: 9188-9193

Barns, S.M., Fundyga, R.E., Jeffries, M.W,. & Pace, N.R. (1994) Remarkable archaeal diversity detected in Yellowstone National Park hot spring enviroment. *Proc. Natl. Acad. Sci. USA.* **91**: 1609-1613

Baross, J.A. & Hoffman, S.E. (1985) Submarine hydrothermal vent and associated gradient environment as sites for the origin and evolution of life. *Origins of Life.* **15**: 327-345

Benachenhou-Lahfa, N., Forterre, P. & Labedan, B. (1993) Evolution of glutamate dehydrogenase genes: Evidence for two paralogous protein families and unusual branching patterns of the archaebacteria in the universal tree of life. *J. Mol. Evol.* **36**: 335-346

Bintrim S.B., Donohoe T.J. Handelsman J., Roberts G.P. & Goodman R.M. (1997) Molecular phylogeny of Archaea from soil *Proc. Natl .Acad. Sci. USA.,* **94:** 277-282

Bischoff, K.M. & Rodwell, V.M. (1996) 3-Hydroxyl-3-methylglutaryl-coenzyme A reductase from *Haloferax volcanii* Purification, characterisation, and expression in *Escherichia coli. J. Bact.* **178**: 19-23

Blecher, O., Goldman, S. & Mevarech, M. (1993) High expression in *Esherichia coli* of the gene coding for the dihydrofolate reductase of the extremely halophilic archaebacterium *Haloferax volcanii* – Reconstitution of the active enzyme and mutation studies. *Eur. J. Biochem.* **216**: 199-203

Bohm S.B & Jaenicke R. (1994) A structure based model for the halophilic adaptation of dihydrofolate-reductase from *Halobacterium-volcanii. Prot. Eng.* **7**: 213-220

Bonelo, G., Veentoa, A., Megias, M. & Ruiz-Berraquero, F. (1984) The sensitivity of halobacteria to antibiotics. *FEMS Microbiol. Lett.* **21**: 341-345

Bonete, M.L., Llorca, F.I. & Cadenas, E. (1992) Alkaline para-nitrophenylphosphate phosphatase-activity from *Halobacterium-halobium* selective activation by manganese and effect. *J. Biochem.* **245**: 839-845

Borneman, J. & Triplet, E.W. (1997) Molecular microbial diversity in soils from Eastern Amazonia: evidence for unusual microorganisms and microbial population shifts associated with deforestation. *Appl. Environ. Microbiol.* **63**: 2647-2653

Borowitzka, L.J.& Brown, A.D. (1974) The salt relations of marine and halophilic species of the unicellular green algae *Dunaliella.* The role of glycerol as a compatible solute. *Arch. Microbiol.* **96**: 37

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilising the principle of protein-dye binding. *Anal. Biochem.* **72**: 248-254

Brown, A.D. & Cho, K.Y. (1970) The walls of the extremely halophilic cocci: Gram-positive bacteria lacking muramic acid. *J. Gen. Microbiol.,* **62**: 267-270

Brown, J.R. & Doolittle, W.F. (1995) Root of the Universal tree of life based on ancient aminoacyl-transfer-RNA synthetase gene duplications. *Proc. Natl. Acad. Sci. USA.***92:** 2441-2445

Brown, J.R., Masuchi, Y., Robb, F.T. & Doolittle, W.F. (1994) Evolutionary relationships of bacterial and archaeal glutamine synthetase genes. *J. Mol. Evol.* **38:** 566-576

Brown, J.W., Daniels, C.J. & Reeve, J.N. (1989) Gene structure, organisation and expression in archaebacteria. *CRC Critl Rev Microbiol.* **16**: 287-338

Bult, C.J., White, O., Olsen, G.J., Zhou, L.X., Fleischmann, R.D., Sutton, G.G., Blake, J.A., FitzGerald, L.M., Clayton, R.A., Gocayne, J.D., Kerlavage, E.F., Weinstock, K.G., Merrick, J.M., Glodek, A., Scott, J.L., Geoghagen, N.S.M., Weidman, W.F., Fuhrmann, J.L., Nguyen, D., Utterback, T.R., Kelly, J.M., Peterson, J.D., Sadow, P.W., Hanna, M.C., Cotton, M.D., Roberts, K.m., Hurst, M.A., Kaine, B.P., Borodovsky, M., Klenk, H.P., Fraser, C.M., Smith, H.O., Woese, C.R. & Venter. J.C. (1996) Complete genome sequence of the methanogentic archaeon *Methanococcus jannaschii. Science.* **273** : 1058-1073

Cendrin, F., Chroboczek, J., Zaccai, G., Eisenburg, H. & Mevarech, M. (1993) Cloning, sequencing and expression on *Escherichia coli* of the gene coding for malate dehydrogenase of the extremely halophilic archaebacterium *Haloarcula marismartui. Biochem,* **32**: 4308-4313

Charlebois, R.L., Hopman, J.D., Schalkwyk, L.C., Lam, W.L. & Doolittle, W.F. (1989) Genome mapping in halobacteria. *Can. J. Microbiol.* **35**: 21-29

Charlebois, R.L., Lam, W.L., Cline, S.W., & Doolittle, W.F. (1987) Characterisation of phv2 from *Halobacterium volcanii* and its use in demonstrating transformation of an archaebacterium. *Proc. Natl. Acad. Sci. USA.*, **84**, 8530-8534

Charlesbois, R.L., Schalkwyk, L.C., Hoffman, J.D. & Doolittle, W.R. (1991) Detailed physical map and set of overlapping clones covering the genome of the archaebacterium *Haloferax volcanii* DS2. *J. Mol. Biol.* **222**: 509-524

Connaris, H., Chaudhuri, J.B., Danson, M.J. & Hough, D.W. (1999) Expression, reactivation and purification of enzymes from *Haloferax volcanii* in *Escherichia coli. Biotech. Bioeng.* **64**: 38-45

Conover, R.K. & Doolittle, W.F. (1990) Characterisation of a gene involved in histidine biosynthesis in Halobacterium *Haloferax volcanii* isolation and rapid mapping by transformation of an auxotroph with cosmid DNA. *J Bact,* **172**: 3244-3249

Creti, R., Londei, P. & Cammarano, P. (1993) Complete nucleotide sequence of an archaeal (*Pyrococcus woesei*) gene encoding a homolog of eukaryotic transcription factor IIB (TFIIB) *Nucl. Acids Res.* **21**: 2942-2942

Dane, M., Steinert, K., Esser, K., Bicklesandkotter, S. & Rodriguez-Valera, F. (1992) Properties of the plasma-membrane atpases of the halophilic archaebacteria *Hlaloferrax volcanii* and *Haloferax mediterranei. Zeitschrift fur naturforshung Can. J. Biosci.* **47**: 835-844

Danner, S. & Soppa, J. (1996) Characterisation of the distal promoter element of halobacteria in *vivo* using saturation mutagensis and selection. *Mol. Micrbiol.* **19**: 1265-1276.

Danson M.J. & Hough, D.W. (1997) The Structural Basis of Protein Halophilicity *Comp. Biochem. Physiol.* **117A**: Elsevier Science Inc.

Danson, M,J., Hough, D.W,. Lunt, G.G (eds.) (1992) The Archaebacteria: biochemistry and biotechnology. Portland press, London

Danson, M.J. (1988) Dilhydrolipoamide dehydrogenase: a 'new' function for an old enzyme? *Biochem. Soc. Trans.* **16**: 87-89

Danson, M.J. (1993) Central metabolism of the Archaea. *New Compr. Biochem.*, **26**: 1-24

Danson, M.J., Black, S.C., Woodland, D.L. & Wood, P.A. (1985) Citric-acid cycle enzymes of the archaebacteria citrate synthase and succinate thiokinase. *FEBS. Microbiol lett.* **179**: 120-124

Danson, M.J., McQuattie, A. & Stevenson, K.J. (1986) Dihydrolipoamide dehydrogenase from halophilic archaebacteria. Purification and properties of the enzyme from *Halobacterium-halobium. Biochem* **23**: 3880-3884

De Medicis, E., Laliberte, J.F. & Vassmarengo, J. (1982) Purification of pyruvate-kinase from *Halobacterium-cutirubrum. Biochm et biophys acta.* **708**: 57-67

Delong, E.F. (1992) Archaea in coastal marine enviroments. *Proc. Natl. Acad. USA.* **89**: 5685-5689

Delong, E.F. (1994) Antarctic microbiology. *Science,* **263**: 401-402

Dennis, P.P. & Shimmin, L.C. (1997) Evolutionary Divergence and Salinity-Mediated Selection in Halophilic Archaea. *Microbiol and Molbiol Rev.* **61**: 90-104

Doolittle, W.F. & Brown, J.R. (1994) Tempo, mode, the progenote, and the universal root. *Proc. Natl. Acad. Sci. USA,* **91**: 6721-6728

Dundas, I.E.D (1977) Physiology of Halobacteria. *Adv. Microb. Physiol.* **15**: 85-120

Dyall-Smith M., Holmes M., Nuttal S. & Woods W. (1995) Protocols for halobacterial genetics version 2.3

Dym, O., Mesarech, M. & Sussman, J.L. (1995) Structural features that stabilize halophilic malate dehydrogenase from archaebacterium. *Science.* **267**: 1344-1346

Ebel, C., Guinet, F., Langowski, J., Urbanke, C., Gagnon, J. & Zaccai, G. (1992) Solution studies of the elongation factor Tu from the extreme halophilic *Halobacterium marismortui. J. Mol. Biol.* **223**: 361-371

Eisenburg, H. & Wachtel, E.J. (1987) Structural studies of halophilic proteins, ribosomes and organelles of bacteria adapted to extreme salt concentrations. *Annu. Rev. Biophys. Chem.*, **16**: 69-92

Eisenburg, H., Mesarech, M. & Zaccai, G. (1992) Biochemical, structural and molecular genetic aspects of halohilism. *Adv, Prot. Chem.* **43**: 1-62

Eisenthal, R. & Cornish-Bowden, A. (1974) The direct linear plot. A new graphical procedure for estimating enzyme kinetic parameters. *Biochem. J.* **139**: 715-720

Elcock, E.H. & McCammon, J.A. (1998) Electrostatic contributions to the stability of halophilic proteins. *J. Mol. Biol.* **280**: 731-748

Else, A.J. Danson, M.J. & Weitzman, P.D.J.(1988) Models of proteolysis of oligomeric enzymes and their applications to the trypsinolysis of citrate synthases. *Biol. J.* **254**: 437-442

Ermler, U., Merckel, M.C., Thauer, R.K. & Shima, S. (1997) Formylmethanofuran: tetrahydromethanopterin formyltransferase from Methanopyrus kandleri-new insights into salt-dependence and thermostability. *Structure,* **5**: 635-646

Feinberg A.P. & Vogelstein B. (1984) " A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity." Addendum. *Anal Biochem* **137:** 266

Feinberg, A.P. & Vogelstein, B. (1983) A technique for radiolabeling DNA restriction endonuclease fragments to specific activity. *Anal Biochem.,* **132**: 6-13

Felsenstein, J. (1993) PHYLIP (Phylogeny Inference Package), *University of Washington, Seattle, USA.*

Fewson, C.A. (1986) Archaebacteria. *Biochem. Edu.* **14**: 103-115

Forterre, P. (1996) A Hot Topic: The Origin of Hyperthermophiles. *Cell.* **85**: 789-792

Forterre, P., Benachenhou-Lahfa, N., Confalonieri, F., Duguet, M., Elie, C. & Labedan, B. (1993) The nature of the last universal ancestor and the root of the tree of life, still open questions. *Biosystems.* **28**: 15-32

Frey, G., Thomm, M., Brudigam, H., Gohl, P. & Hausner, W. (1990) An Archaebacterial cell-free transcription system: the expression of tRNA genes from *Methanococcus vannielii* is mediated by a transcription factor. *Nuc. Acid. Res.* **18**: 1361-1367

Frolow, F., Harel, M., Sussman, J.L., Mevarech, M. & Shoham, M. (1996) Insights into protein adaptation to a saturated salt environment from the crystal struture of a halophilic 2Fe-2S ferredoxin. *Nat. Struct. Biol.* **3:** *5*

Fuhrman J.A.K. (1992) Novel marine archaebacterial group from marine plankton *Nature* **356:** 148-149

Fukumori, Y., Fujiwara, T., Okadatakahashi, Y., Mukohata, Y. & Yamanaka, T. (1985) Purification and properties of a peroxidase from *Halobacterium halobium* L 33. J. *Biochem.* **98**: 1055-1061

Gerike, U., Danson, M.J., Russell, N.J. & Hough, D.W. (1997) Sequencing and expression of the gene encoding a cold-active citrate synthase from an Antarctic bacterium, strain DS2-3R. *Eur. J. Biochem.* **248**: 49-57

Gerike, U., Russell, R.J.M., Danson, M.J., Russell, N.J., Hough, D.W. & Taylor, G.L. (1998) Preliminary crystallographic studies of citrate synthase from an Antarctic psychrotolerant bacterium. *Acta Crystallographica section D-Biological crystallograhpy.* **54**: 1012-1013

Gogarten, J.P., Kibak, H., Dittrich, P., Taiz, L., Bowman, E.J., Bowman, B.J., Manolson, M,F., Poole, R.J., Date, T., Oshima, T., Konishi, J., Denda, K. & Yoshida, M. (1989) Evolution of the vacuolar H+-ATPase: Implications for the origin of eukaryotes. *Proc. Natl. Acad. Sci. USA,* **86**: 6661-6665

Gogarten-Boekels, M., Hilario, E. & Gogarten J.P. (1995) The effects of heavy bombardment on the evolution-The emergence of the three domains of life. *Origins of Life,* **25**: 251-264

Good, W.A. & Hartman, P.A. (1970) Properties of the amylase from *Halobacterium halobium. J. Bact.* **104:** 601-603

Goodford, P.J. (1985) A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **28:** 849-857

Gouy, M. & Li, W.H. (1989) Phlogenetic analysis based on rRNA sequences supports the archaebacterial rather than the eocyte tree. *Nature,* **339**: 145-147

Grant. W.D. & Larsen. H. (1989) Extremely halophilic archaeobacteria. Order Halobacteriales red. nov. In *Bergey's manual of systematic bacteriology* (ed. N. Pfennig), pp.2216-2233. Williams and Wilkins, Baltimore

Gupta, R.S. & Singh, B. (1994) Phylogenetic analysis of 70 kD heat shock protein sequences suggests a chimeric origin for the eukaryotic cell nucleus. *Curr Biol* **4**: 1104-1114

Gutierrez, M.C., Garcia, M.T., Ventosa, A., Nieto, J.J. & Ruizberraquero, F. (1986) Occurrence of megaplasmids in halobacteria. *J. Appl. Bact.* **61:** 67-71

Gutierrez, M.C., Ventosa, A. & Ruizberraquero, F. (1989) DNA-DNA homology studies among strains of \Haloferax and other halobacteria. Curr. Microbiol. 18: 253-256

Gutierrez, M.C., Ventosa, A. & Ruizberraquero, F. (1990) deoxyribonucleic acid reletedness among species of *Haloarcula* and other halobacteria. *Biochem. Cell. Biol.* **68:** 106-110

Hain, J., Reiter, W.D., Hudepohl, U. & Zillig,W. (1992) Elements of an archaeal promoter defined by mutational analysis. *Nuc Acid Res.* **20**: 5423-5328

Hausner, W. & Thomm, M.(1993) Purification and characterisation of a general transcription factor, aTFB from the archaeon *Methanococcus thermlithotrophicus. J. Biol. Chem.* **268**: 24047-24052

Hershberger, K.L., Barns, S.M., Reysenbach, A.L., Dawson, S.C. & Pace, N.R. (1996) Wide diversity of Crenarchaeota. *Nature.* **384**: 420

Holland, H.D. (1984) The Chemical Evolution of the Atmosphere and Oceans. Princeton University Press, Princeton, NJ. 582 pp.

Holmes, M.L., Nuttall, S.D. & Dyall-Smith, M.L. (1991) Construction and use of halobacterial shuttle vectors and further studies on *Haloferax* DNA gyrase. *J. Bact,* **173**: 3807-3813

Holmes, M.L., Pfeifer, F. & Dyall-Smith, M.L. (1994) Improved shuttle vectors for *Haloferax volcanii* including a dual-resistance plasmid. Gene, **146**: 117-121

Holmes, P.K., Dundas, I.E. & Halvorson, H.O. (1965) Halophilic enzymes in cell-free extracts of *Halobacterium salinarium. J. Bact.* **90:** 1159-1160

Hori, H. & Osawa, S. (1986) Evolutionary change in 5S rRNA secondary structure and a phylogenetic tree of 352 5S rRNA species. *Biosys.* **19:** 163-172

Hudepohl, H., Reiter, W.D. & Zillig, W. (1990) In vitro transcription of two rRNA genes of the archaebacterium *Sulfolobus* sp. B12 indicates a factor requirement for specific intiation. *Proc. Natl. Acad. Sci. USA.* **87**: 5851-5855

Iwabe, N., Kuma, k., Hasegawa, M., Osawa, S. & Miyata, T. (1989) Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc. Natl. Acad. Sci.* USA, **86**: 9355-9359

Jaenicke, R. (1991) Protein stability and molecular adaptation to extreme conditions. *Eur. J. Biochem.* **202**: 715-728

James K.D. (1994) Cloning of citrate synthase from the halophilic archaeon *Haloferax volcanii* PhD thesis, University of Bath

James, K.D., Russell, R.J.M., Parker, L., Daniel, R.M., Hough, D.W. & Danson, M.J. (1994) Citrate synthases from the archaea – devlopment of a bio-specific affinity-chromatography purification procedure. *FEMS. Microbiol Lett.* **119**: 181-185

Jolley, K.A., Rapaport, E., Hough, D.W., Danson, M.J., Woods, W.G. & Dyall-Smith, M.L. (1996) Dihydrolipoamide dehdrogenase from the halophilic archeaon *Haloferax volcanii* – Homologous over expression of the cloned gene. *J. Bact.* **178**: 3044-3048

Jolley, K.A., Russell, R.J.M., Hough, D.W. & Danson, M.J. (1997) Site-directed mutagenesis and halophilicity of dihydrolipoamide dehydrogenase from the halophilic archaeon, *Haloferax volcanii. Euro. J. Biochem*, **248**: 362-368

Jones, W.J., Nagle, D.P., Jr., Whitman, W.B. (1987) Methanogens and the diversity of archaebacteria. Microbiol. Rev., **51**: 135-177

Jurgens, G. & Saano, A. (1999) Diversity of soil Archaea in boreal forest before, and after clear-cutting and perscribed burning. *FEMS. Microbiol Ecol.* **29**: 205-213

Jurgens, G., Lindström, K. & Saano, A. (1997) Novel group within the kingdom Crenarchaeota from boreal forest soil. *App. Environ. Microbiol.* **63**: 803-805

Kabsch & Sander (1983) Dictionary of protein secondary structure pattern recognition of hydrogen bonded and geometrical features. *Biopol.* **22:** 2577-2637

Kates, M. (1992) Archaebacterial lipids: Structure, biosynthesis and function. *Biochem. Soc. Symp.* **58**: 51-72
Kawarabayasi Y, Hino Y, Horikawa H, Yamazaki S, Haikawa Y, Jin-no K, Takahashi M, Sekine M, Baba S, Ankai A, Kosugi H, Hosoyama A, Fukui S, Nagai Y, Nishijima K, Nakazawa H, Takamiya M, Masuda S, Funahashi T, Tanaka T, Kudoh Y, Yamazaki J, Kushida N, Oguchi A, & Kikuchi H. (1999) Complete genome sequence of an aerobic hyper-thermophilic crenarchaeon, *Aeropyrum pernix* K1. *DNA Res.* **6**: 145-52

Kawarabayasi Y, Sawada M, Horikawa H, Haikawa Y, Hino Y, Yamamoto S, Sekine M, Baba S, Kosugi H, Hosoyama A, Nagai Y, Sakai M, Ogura K, Otsuka R, Nakazawa H, Takamiya M, Ohfuku Y, Funahashi T, Tanaka T, Kudoh Y, Yamazaki J, Kushida N, Oguchi A, Aoki K, Kikuchi H. (1998) Complete sequence and gene organisation of the genome of a hyper-thermophilic archaebacterium, *Pyrococcus horikoshii* OT3. *DNA Res.* **5**: 55-76

Keradjopoulos, D. & Holldorf, A.W. (1979) Purification and properties of alanine dehydrogenase from *Halobacterium salinarium. Biochim. Biophys. Acta.* **570:** 1-10

Klenk, H.P., Clayton, R.A., Tomb, J.F., White, O., Nelson, K.E., Ketchum, K.A., Dodson, R.J., Gwinn, M., Hickey, E.K., Peterson, J.D., Richardson, D.L., Kerlavage, A.R., Graham, D.E., Kyrpides, N.C., Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S., Kirkness, E.F., Dougherty, B.A., McKenney, K., Adams, M.D., Loftus, B., Peterson, S., Reich, C.I., McNeil, L.K., Badger, J.H., Glodek, A., Zhou, L.X., Overbeek, R., Gocayne, J.D., Weidman, J.F., McDonald, L., Utterback, T., Cotton, M.D., Spriggs, T., Artiach, P., Kaine, B.P., Sykes, S.M., Sadow, P.W., Daandrea, K.P., Bowman, C., Fujii, C., Garland, S.A., Mason, T.M., Olsen, G.J., Fraser, C.M., Smith, H.O., Woese, C.R. & Venter, J.C. (1997) The complete genome sequence of the hyperthermophilic, sulphate-L reducing archaeon Archaeoglobus fulgidus. Nature. **390**: 364-379

Krishnan, G. & Altekar, W. (1993) Halophilic class-I aldolase and glyceraldehyde-3-phosphate dehydrogenase some salt dependent structural features. *Biochem* **32**: 791-798

Kristjansson, H. & Hochstein, L.I. (1985) Dicyclohexylcarbodiimide–sensitive ATPase in *Halobacterium saccharovorum. Arch. Biochem. Biophys.* **85:** 590-595

Kuchner O.A.F. (1997) Directed evolution of enzyme catalyts *TIBS* **15:** 523-530

Kudo, Y., Shibata, S., Miyaki, T., Aono, T. & Oyaizu, H. (1997) Peculiar Archaea found in Japanese paddy soils. *Biosci. Biotech. Biochem.* **61**: 917-920

Kuntz, I.D. (1971) Hydration of marcomolecules. III. Hydration of polypeptides. *J. Am. Chem. Soc.* **93**: 514-516

Kushner, D.J. (1988) What is the "true" internal environment of halophilic and other bacteria. *Can. J. Microbiol.* **34**: 482-486

Kushner, D.J., Bayley, S.T., Boring, J., Kates, M. & Gibbons, N.E. (1964) Morphological and chemical properties of cell enevlopes of the extreme halophile *Halobacterium cutirubrum. Can. J. Bact,* **173**: 5352-5358

Labedan, B., Forterre, P. & Benachenhou-Lahfa, N. (1993) *J. Mol. Evol.* **36**, 335-346

Lake, J.A., Henderson, E., Oakes, M. & Clark, M.W. (1984) Eocytes - A new ribosome structure indicates a kingdom with a close relationship to eukaryotes. *Proc. Natl. Acad. Sci. USA.* **81:** 3786-3790

Lake, J.A. (1988) Origin of the eukaryotic nucleus determined by rate-invariant analysis of rRNA sequences. *Nature,* **331**: 184-186

Lake, J.A. (1989) Origin of the eukaryotic nucleus: eukaryotes and eocytes are genotypically related. *Can. J. Microbiol.*, **35**: 109-119

Lam, W.L. & Doolittle, W.F. (1989) Shuttle vectors for the archaebacterium *Halobacterium volcanii. Proc. Natl. Acad. Sci.* USA., **86**: 5478-5482

Lam, W.L., Cohen, A., Tsouluhas, D. & Doolittle, W.F. (1990) Genes for tryptophan biosynthesis in the archaeabacterium *Haloferax volcanii. Proc. Natl. Acad. Sci. USA.* **87:** 6614-6618

Langer, D., Hain, J., Thuriaux. & Zilling, W. (1995) Transcription in the archaea similarity to that Eucarya. *Proc. Natl. Acad. Sci. USA* **92**: 5768-5772

Lanyi, J.K. (1974) Salt-dependent properties of proteins from extremely halophilic bacteria. *Bact. Rev.* **38**: 272-290

Lanyi, J.K. & Stevenson, J. (1969) Effect of salts and organic solvents on the activity of *Halobacterium cutirubrum* catalase. *J. Bact.* **98:** 611-616

Louis, B.G. & Fitt, P.S. (1972) Purification and properties of the ribonucleic acid-dependent ribonucleic acid polymerase from *Halobacterium cutirubrum. Biochem. J.* **128:** 755-762

MacGregor, B.J., Moser, D.P., Wheeler Alm, E., Nealson, K.H. & Stahl, D.A. (1997) Crenarchaeota in Lake Michigan sediment. *Appl. Environ. Microbiol.* **63**: 1178-1181

Madigan, M.T., Martinko. J.M. & Parker, J. (1997) Brock Biology of Microorganisms. *8th ed NJ. Prentice hall*, 169-171, 173, 743-747

Margulis, L. & Guerrero, R. (1991) Kingdoms in turmoil. *New Scientist*, 23 March 46-50

Margulis, L. (1996) Archaeal-eubacterial mergers in the origin of Eukarya: Phlogenetic classification of life. *Proc. Natl. Acad. Sci. USA*, **93**: 1071-1076

Marsh, T., Reich, C., Whitelock, R. & Olsen, G. (1994) Transcription factor IID in the Archaea : Sequence in the Teromococcus celer genome would encode a product closely related to the TATA-binding protein of eukaryotes. *Proc. Natl. Acad. Sci. USA.* **91**: 4180-4184

May, B.P. & Dennis, P.P. (1987) Sueroxide-dismutase from the extremely halophilic archaebacterium *Halobacterium cutirubrum. J. Bacteriol.* **169**: 1417-1422

Mayer, E. (1990) Nature, **348**: 491

Meinkoth, J. & Whal, G. (1984) Hybridization of nucleic-acids immobilised on solid supports. *Anal. Biochem.*, **138**: 267-284

Mevarech, M. & Werczberger, R. (1985) Genetic transfer in Halobacterium-volcanii. *J. Bact*, **162**: 461-462

Mevarech, M., Eisenberg, H. & Neumann, E. (1977) Malate dehydrogenase isolated from extremely halophilic bacteria of the Dead Sea. 1. Purification and molecular characterisation. *Biochem.*, **16**: 3781-3785

Mullakhanbhai M.F. & Larsen (1975) Halobacterium volcanii spec. nov., a Dead Sea halobacerium with a moderate salt requirement. *Arch. Microbiol.* **104:** 207-214

Nanba, T. & Mukohata, Y. (1987) A membrane-bound atpase from Halobacterium-halobium purification and characterisation. J. *Biochem.* **102**: 591-598

Nicholls, A., Bharadwaj, R. & Honig, B. (1993) GRASP-Graphical representation and analysis of the surface properties. *Biophysical J.*, **64**: 166

Norberg, P., Kaplan, J.G. & Kushner, D.J. (1973) Kinetics and regulation of the salt-dependent aspartate transcarbamylase of *Halobacterium cutirubrum. J. Bact.* **113:** 680-686

Oesterhelt, D. & Tittor, J. (1989) Two pumps, one principle: Light-driven ion transport in Halobacteria. *TIBS* **14**: 57-61

Olson, G.J. & Woese, C.R. (1993) Ribosomal RNA: a key to phylogeny. *FASEB J.*, **7**: 113-123

Oren, A. (1983) Halobacterium-sodomense SP-NOV, A dead sea Halobacterium with an extremely high magnesium requirement. *Int. J. Syst. Bact* **33**: 381-386

Ouzounis, C. & Sander, C.(1992) TFIIB, an evolutionary link between the transcription machineries of archaebacteria and eukaryotes. *Cell.* **71**: 189-190

Ovreås, L., Forney, L., Daae, F.L. & Torsvik, V. (1997) Distribution of bacterioplankton in meromictic lake Saelenvannet, as determined by naturing gradient gel-electrophoresis of PCR-amplified gene fragments coding for 16S rRNA. *Appl. Environ. Microbiol.* **63**: 3367-3373

Pace, N.R. (1991) Origin of life-facing up to the physical setting. *Cell*, **65**: 531-533

Pace, N.R. (1996) New perspective on the natural microbial world: Molecular microbial ecology. *ASM News.* **62:** 463-470

Palmer, J.R. & Daniels, C.J. (1995) In vivo definition of an archaeal promoter. *J. Bact.* **177**: 1844-1849

Pfeifer, F., Weidinger, G. & Goebel, W. (1981) Genetic variability in *Halobacterium halobium. J. Bact.* **145:** 375-381

Pieper, U., Brinkmann, T., Kruger, T., Noyerweidner, M, & Pingoud, A. (1997) Characterisation of the interaction between the restriction endonuclease McrBC from *E.coil* and its cofactor GTP. *J. Mol. Biol.* **272**: 190-199

Preston, G.M., Wu, k.y., Molinski, T.F., Delong, E.F. (1996) A psychrophilic crenarchaeon inhabits amarine sponge – crenarchaeum symbiosum gen-nov., sp., nov. *Proc. Natl. Acad. Sci. USA*, **93**: 6241-6246

Qureshi, S., Baumann, P., Rowlands, T., Khoo, B. & Jackson, S. (1995) Cloning and functional analysis of the TATA binding protein from *sulfolobus shibatae. Nuc Acid Res.* **23**: 1775-1781

Rao, J.K. & Argos, P. (1981) Structural stability of halophilic proteins. *Biochem.* **20**: 6536-6543

Reistad, R. (1970) On the composition and nature of the bulk protein of extremely halophilic bacteria. *Arch. Microbiol.*, **71**: 353-360

Reiter, W.D., Palm, P. & Zillig, W. (1988) Analysis of transcription in the archbacterium *Sulfolobus* indicates that archaebaterial promoters ara homologous to eucaryotic pol II promoters. *Nucl Acid Res.* **16**: 1-19

Remington, S., Wiegand, G. & Huber, R. (1982) Crystallographic refinement and atomic models of2 different forms of citrate synthase at 2.7-A and 1.7-A resolution. *J. Mol. Biol.* **158**: 111-152

Robb, F.T., Place, A.R. & Savers, K.R. (1995) Archaea: A Laboratory Manual. Cold Harbour. *Labatory Press*, 3-13

Rodriguez-Valera F., Ruiz-Berraquero F. & Ramos-Cormenzana (1980) Isolation of extremely halophilic bacteria able to grow in defined inorganic media with single carbon sources. *J. Gen. Microbiol.* **119:** 535-538

Rodriguez-Valera, F.(1988) Characteristics and microbial ecology of hypersaline environments. In *Halophilic* bacteria (Rodriguez-Valera, F., ed.) Vol 1, pp 3-30. CRC Press

Rodriguez-Valera, F., Juez, G. & Kushner, D.J. (1983) *Halobacterium mediterranei* spec. nov., a new carbonhydrate-utilizing exteme halophile. *Sys. Appl. Microbiol.* **4**: 369-381

Rowlands, T., Baumann, P. & Jackson, P. (1994) The TATA-binding protein: A general transcription factor in Euraryotes and Archaebacteria. *Science.* **264**: 1326-1329

Russell, R.J.M., Ferguson, J.M.C., Hough, D.W., Danson, M.J. & Taylor, G.L. (1997) The crystal structure or citrate synhaese from the hyperthermophilic Archaeon pyrococcus furiosus at 1.9 angstrom resolution. *Biochem.* **36**: 9983-9994

Russell, R.J.M., Gerike, U., Danson, M.J., Hough, D.W. & Taylor, G.L. (1998) Structural adaptations of the cold-active citrate synthase from an Antarctic bacterium. *Structure*, **6**: 351-361

Russell, R.J.M., Hough, D.W., Danson, M.J. & Taylor, G.L. (1994) The crystal structure of citrate synthase from thermophilic archaeon, *Thermoplasma acidophilum. Structure.* **12**: 1157-1167

Ryu, K.G., Kim, J.B. & Dordick, J.S. (1994) Catalytic properties and potential of a extracellular protease from an extreme halophile. *Enz Microb Tech.* **16**: 266-275

Salin, M.L. & Oestrhelt, D. (1988) Purification of a manganese containing superoxide-dismutase from *Halobacterium-halobium. Arch Biochem. Biophys.*, .**260**: 806-810

Sambrook, J., Maniatis, T. & Fritsch, E.F. (1989) Molecular cloning, a laboratory manual. Cold Spring Harbour Laboratory, Cold Spring Harbour, N.Y.

Schleper, C., Holben, W. & Klenk, H.P. (1997) Recovery of Crenarchaeotal ribosomal DNA sequences from freshwater-lake sediments. *Appl. Environ. Microbiol.* **63**: 321-323

Seedgouse C. (1995) The cloning and sequencing of citrate synthase from the halophilic archaea *Haloferax volcanii* Final year project, University of Bath

Shima, S., Weiss, D.S. & Thauer, R.K. (1995) Formylmethanofuran – Tetrahydromethanopterin Formyltransferase (FTR) from the hyperthermophilic methhanpyrus-Kandleri – cloning sequencing and functional expression of the FTR gene and one-step purification of the enzyme overproduced in *Escherichia coli. Eur. J. Biochem.* **230**: 906-913

Smith,D.R., DoucetteStamm,L.A., Deloughery,C., Lee,H.M., Dubois,J., Aldredge,T.,Bashirzadeh,R., Blakely,D., Cook,R., Gilbert,K., Harrison,D., Hoang,L., Keagle,P., Lumm, W., Pothier,B., Qiu,D.Y., Spadafora,R., Vicaire,R., Wang,Y., Wierzbowski,J., Gibson,R., Jiwani,N., Caruso,A., Bush,D., Safer,H., Patwell,D., Prabhakar,S., McDougall,S., Shimer,G., Goyal,A., Pietrokovski,S., Church,G.M., Daniels,C.J., Mao,J.I., Rice,P.,Nolling.J., Reeve.J.N. (1997) Complete genome sequence of Methanobacterium thermoautotrophicum delta H: Functional analysis and comparative genomics. *J. Bact.* **179:** 7135-7155

Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoretic. *J. Mol. Biol.* **98**: 503

Spudich. J.L. (1993) Colour sensing in the Archaea: A eukaryotic-like receptor coupled to a prokaryotic transducer. *J. Bact.* **175**: 7755-7761

Srere, P.A., Brazil, H. & Gonen, L. (1963) The citrate condensing enzyme of pigeon breast muscle and moth flight muscle. *Acta. Chem. Scand.* **17**: S129-S134

St. Jean A. & Charlebois, R.L. (1996) Comparative genomic analysis of the *Haloferax volcanii* DS2 and *Halobacterium salinarium* GRB contig maps reveals extensive rearrangement. *J. Bact.* 96: 3860-3868

Stoeckenius, W. (1978) Speculations about the evolution of halobacteria and of the chemiosmotic mechanisms. In *Light transducing memdrane structure, function, and evolution* (ed. D.W Deamer), pp 122-139. Academic Press, New York

Studier F.W. Rosenberg A.H. Dunn J.J. & Dubendorff J.W. (1990) *Meth. Enzymol.* **185:** 60-89

Tabita, F.R. & McFadden (1976) Molecular and catalytic properties of ribulose 1,5-bisphosphate carboxylase from the photosynthetic extreme halophile *Ectothiorhodospira halophila. J. Bact.* **126:** 1271-1277

Taupin, C.M.J., Hartlien, M. & Leberman, R. (1997) Seryl-tRNA synthetase from the extreme halophilic *Haloarcla marismortui* – Isolation, characterisation and sequencing of the gene and its expression in *Escherichia coli*. *Eur. J. Biochem.* **247**: 141-150

Thanki, N., Thornton, J.M. & Goodfellow, J.M. (1988) Distributions of water around amino-acids residues in proteins. *J. Mol. Biol.* **202**: 637-657

Thomm, M. & Wich, G. (1988) An archaebacterial promoter element for stable RNA genes with homology to the TATA box of higher eukaryotes. *NuciAcid Res.* **16**: 151-163

Timasheff, S.N. (1992) Solvent effects on protein stability. *Corr. Op. Struct. Biol.,* **2**: 35-39

Tindall, B.J., Soliman, G.S.H. & Truper, H.G. (1984) Polar lipids of some new archaebacterial halophiles. *Sys. App. Microbiol.* **5**: 275

Tomlinson, G.A., Koch, T.K. & Hochstein, L.I. (1974) The metabolism of carbohydrates by extremely halophilic bacteria: glucose metabolism via a modified Enter Doudoroff pathway. *Can. J. Microbiol,* **20**: 1085-1091

Tornabene, T.G. & Langworthy, T.A. (1979) Diphytanil and dibiphytanil glycerol ether lipids of methanogenic archaebacteria. *Science,* **203**: 51-53

Torreblanca, M., Rodriguez-Valera, F., Juez, G., Ventosa, A., Kamekura, M. & Kates, M. (1986) Classification of non-alkaliphilic halobacteria based on numerical taxonomy and polar lipid-composition, and description of *Halorcula* gen-nov and *Haloferax* gen-nov. *Sys. App. Microbiol.* Vol. **8**, No. **1-2**: 89-99

Ueda, T., Suga, Y. & Matsuguchi, T. (1995) Molecular phylogenetic anaylsis of a soil microbial community in a soy bean field. *Eur. J. Soil. Sci.* **44**: 415-421

Unwins P.J.R., Webb R.I. & Taylor A.P. (1998) Novel nano-organisms from Australian sandstones *Am. Min.* **83:** 1541-1550

Vettakkorumakankav, N.N., Danson, M.J., Hough, D.W., Stevenson, K.J., Davidson, M. & Young, J. (1992) Dilydrolipoamide dehydrogenase from the halophilic archaebacterium *Haloferax volcanii.* characterisation and N-terminal sequence. *Biochem. Cell Biol.*, **70**: 656-663

Wachtershauser, G. (1988) Pyrite formation, the first energy source for life: a hypothesis. *System. Appl. Microbiol.* **10**: 207-210

Wachtershauser, G. (1990) Evolution of the first metabolic cycles. *Proc. Natl. Acad. Sci. USA.* **87**: 200-204

Werber, M.M. & Mevarech, M. (1978) Purification and characterisation of a highly acidic 2Fe-ferredoxin from Halobacterium of the Dead Sea. *Arch. Biochem. Biophys.* **187**: 447-456

Whal, G.M., Stern, M. & Stark, G.R. (1979) Efficient transfer of a large DNA fragments from agaose gels to diazobenzyloxymethyl-paper and rapid hybridization by using dextran sulfate. *Proc. Natl. Acad. Sci.* **76**: 3683

Wheelis, M.L., Kandler, O. & Woese, C.R. (1992) On the nature of global classification. *Proc. Natl. Acad. Sci. USA,* **89**: 2930-2934

Williams & Wilkins Eds, Bergey's manual of systematic bacteriology vol1 USA

Woese, C.R & Fox, G.E. (1977) Phlogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc. Natl. Acad. Sci.* USA, **74**: 5088-5090

Woese, C.R. & Olsm, G,J. (1986) Archaebacterial phylogeny: perspectives on urkingdoms. *System. Appl. Microbiol.*, **7**: 161-177

Woese, C.R. (1987) Bacterial Evolution. *Microbiol. Rev.,* **51**: 221-271

Woese, C.R., Kandler, O. & Wheelis, M.L. (1990) Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria and Eucarya. *Proc. Natl. Acad. Sci. USA.,* **87**: 4576-4579

Woese, C.R., Magrum, L.J. & Fox, G.E. (1978) Archaebacteria. *J. Mol. Evol.* **11**: 245-252

Yip K.S.P., Stillman J.J., Lebbink J., deVoss N.W., Robb F.T., Vetriani C., Maeder D. & Rice D.W. (1998) Insights into the molecular basis of thermal stability from the analysis of ion-pair networks in the glutamine dehydrogenase family *Eur. J. Biochem* **255:** 336-346

Zaccai, G., Cendrin, F., Halk, Y., Borochov, N. & Eisenberg, H. (1989) Stabilization of halophilic malate dehydrogenase. *J. Mol. Biol.* **208:** 491-500

Zillig, W., Klenk, H.P., Palm, P., Puhler, G., Gropp, F., Garrett, R.A. & Leffers, H. (1989) The phylogenetic relations of DNA-dependent RNA-polymerases of archaebacteria, eukaryotes, and eubacteria. *Can. J. Microbiol.* **35**: 73-80

Zillig, W., Palm, P., Klenk, H.P., Langer, D., Hudepohl, U. & Hain, M (1993) Transcription in the Archaea. *Biochem Arch.* **26**: 367-391

Zillig, W., Stetter, K.O. & Tobun, M. (1978) DNA-dependent RNA polymerase *from Halobacterium halobium.* Eur. J. Biochem. **91:** 193-199

Zusman, T., Rosenshine, I., Boehm, G., Jaenicke, R., Leskiw, B. & Mevarech, M. (1989) Dihydrofolate reductase of the extremely Halophilic archaebacterium *Halobacterium volcanii.* the enzyme and its coding gene. *J. Biol. Chem..* **264:** 18878-18883

Appendix 1

Genetic Markers

F' A low copy number self-transmissible plasmid carring portions of the *E.coli* chromosome, most notably the lac operon and ompT .

hsdS

hsdR DNA that does not contain methylation of certain sequences is recognized as foreignby *EcoKI* or *EcoBI* and restricted (degraded). These enzymes recognize different sequences and are encoded by different alleles of *hsdRMS*. *HsdR* mutations aboish restriction but not protective methylation (r^-m^+), while *hsdS* mutations abolish both (r^- m^-). DNA made in the latter will be restricted when introduced into a wild-type strain.

dcm Endogenous cytosine methylation at CCWGG sequences abolished. Used for making DNA susceptible to cleavage by some restriction enzymes (e.g. *Ava II*).

endA Activity of nonspecific Edonuclease I abolished. DNA preparations are thought to be of higher quality when prepared from *endA* strains.

RecA Homologous recombination abolished; particularly desirable when working with sequences containing direct repeats >50 bp.

(P2) the cell carries a P2 prophage. This allows selection against Red^+ Gam^+ λ (Spi-selection).

supE Strains carry a glutamine-inserting amber (UAG) suppressor. Nonsense mutations in the same strain are supressed only at low temperatures.

mcrA

mcrBC These affect methylcytosine-specific restriction systems. DNA containing methylcytosine in some sequences is restricted by Mcr+. Dcm-modified DNA is not restricted by Mcr^+. *Δ(mcrC-mrr)* deletes six genes; *mcrC-mcrB-hsdS-hsdM-hsdR-mrr, mcrA* is lost with e14.

lacI^q Overproduces the lac repressor, turning off expression from *P lac* more completely.

lacZ β-galactosidase activity abolished.

Δ(lac) = deletion; there are four common deletions involving *lac*.

Δ(lacZ

M15 expresses a fragment that complements the lac α-fragment encoded by many vectors. These vectors will yield blue colour on X-Gal only if the host carries ΔM15.

pro AB The cell does not require proline for growth on minimal medium.

Appendix 2

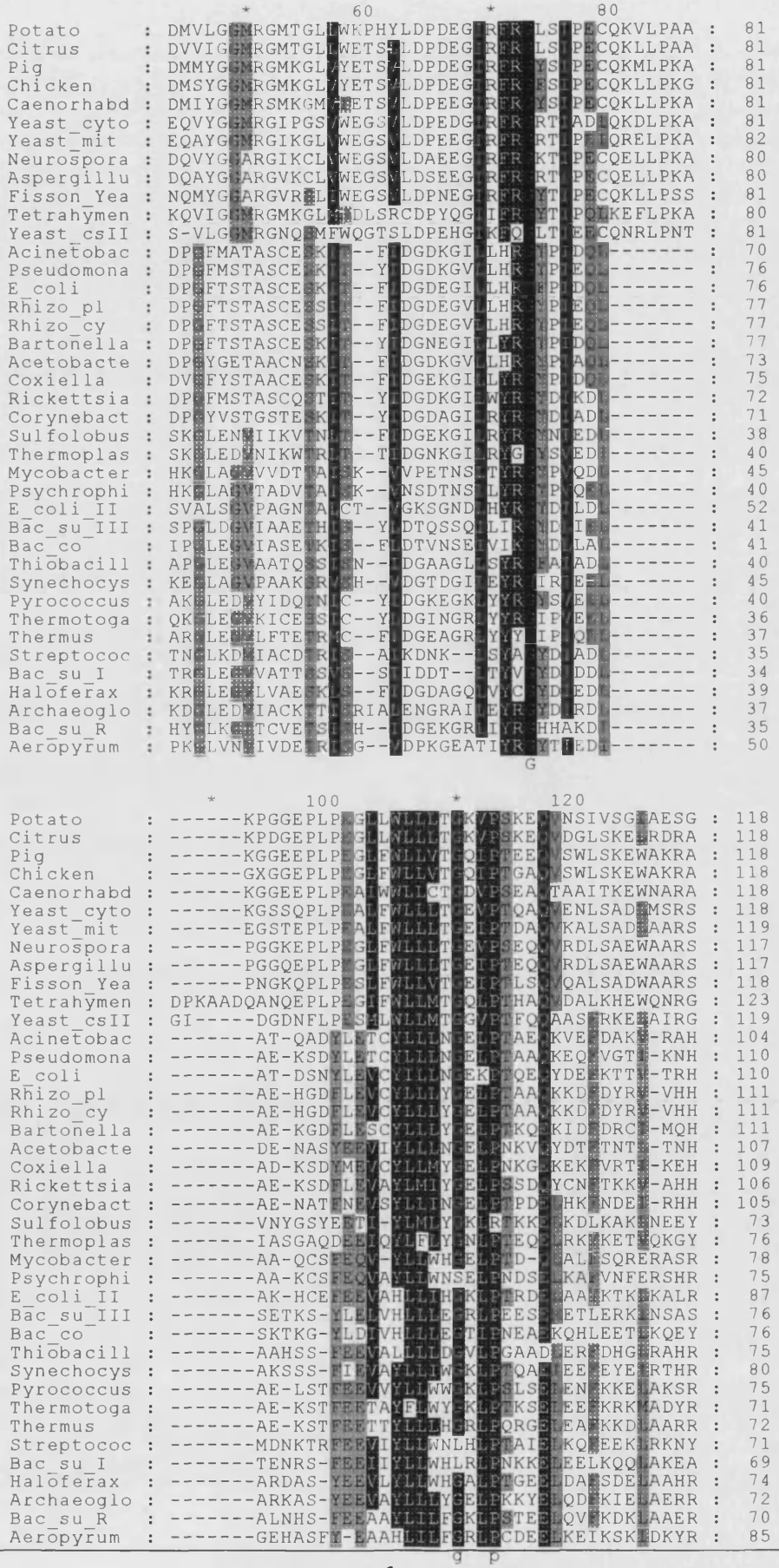
Organism	Abbreviation	Domain	Classification
Haloferax volcanii	Haloferax	Archaea	Euryarchaeota
Pryococcus furiosus	Pyrococcus	Archaea	Euryarchaeota
Sulfolobus sulfataricus	Sulfolobus	Archaea	Crenarchaeota
Thermoplasma acidophilum	Thermoplas	Archaea	Euryarchaeota
Bacillus Coagulans	Bac co	Eubacteria	Gram+ve
Bacillus subtilis (csiR)	Bac su R	Eubacteria	Gram+ve
Bacillus subtills (csi)	Bac su I	Eubacteria	Gram+ve
Bacillus subtills (csiii)	Bac su III	Eubacteria	Gram+ve
Corynebacterium gultamicum	Corynebact	Eubacteria	Gram+ve
Mycobactium smegmatis	Mycobacter	Eubacteria	Gram+ve
Psychrophilic isolate (DS2-3S)	Psychrophi	Eubacteria	Gram+ve
Streptococcus mutans	Streptococ	Eubacteria	Gram+ve
Acetobacter aceti	Acetobacte	Eubacteria	Gram-ve
Acinobacter anitratium	Acinobacte	Eubacteria	Gram-ve
Bartonella henselae	Bartonella	Eubacteria	Gram-ve
Coxiella burnetii	Coxiella	Eubacteria	Gram-ve
Escherichia coli	E. coli	Eubacteria	Gram-ve
Escherichia coli II	E. coli II	Eubacteria	Gram-ve
Rhizobuim tropici (cytoplasmic)	Rhizo cy	Eubacteria	Gram-ve
Rhizobuim tropici (plasmid)	Rhizo pl	Eubacteria	Gram-ve
Synechocystis sp	Synechocys	Eubacteria	Gram-ve
Thermus aquaticus	Thermus	Eubacteria	Gram-ve
Thiobacillus ferroxidans	Thiobacill	Eubacteria	Gram-ve
Aspergillus niger	Aspergillu	Eukaryote	Fungi
Caenorhabditis elegans	Caenorhabd	Eukaryote	Nematode
Citrus masima	Citrus	Eukaryote	Plant
Gallus gallus	Chicken	Eukaryote	Animal
Neurospora erassa	Neurospora	Eukaryote	Fungi
Saccharomyces cerevisiae (cii)	Yeast csII	Eukaryote	Fungi
Saccharomyces	Yeast cyto	Eukaryote	Fungi

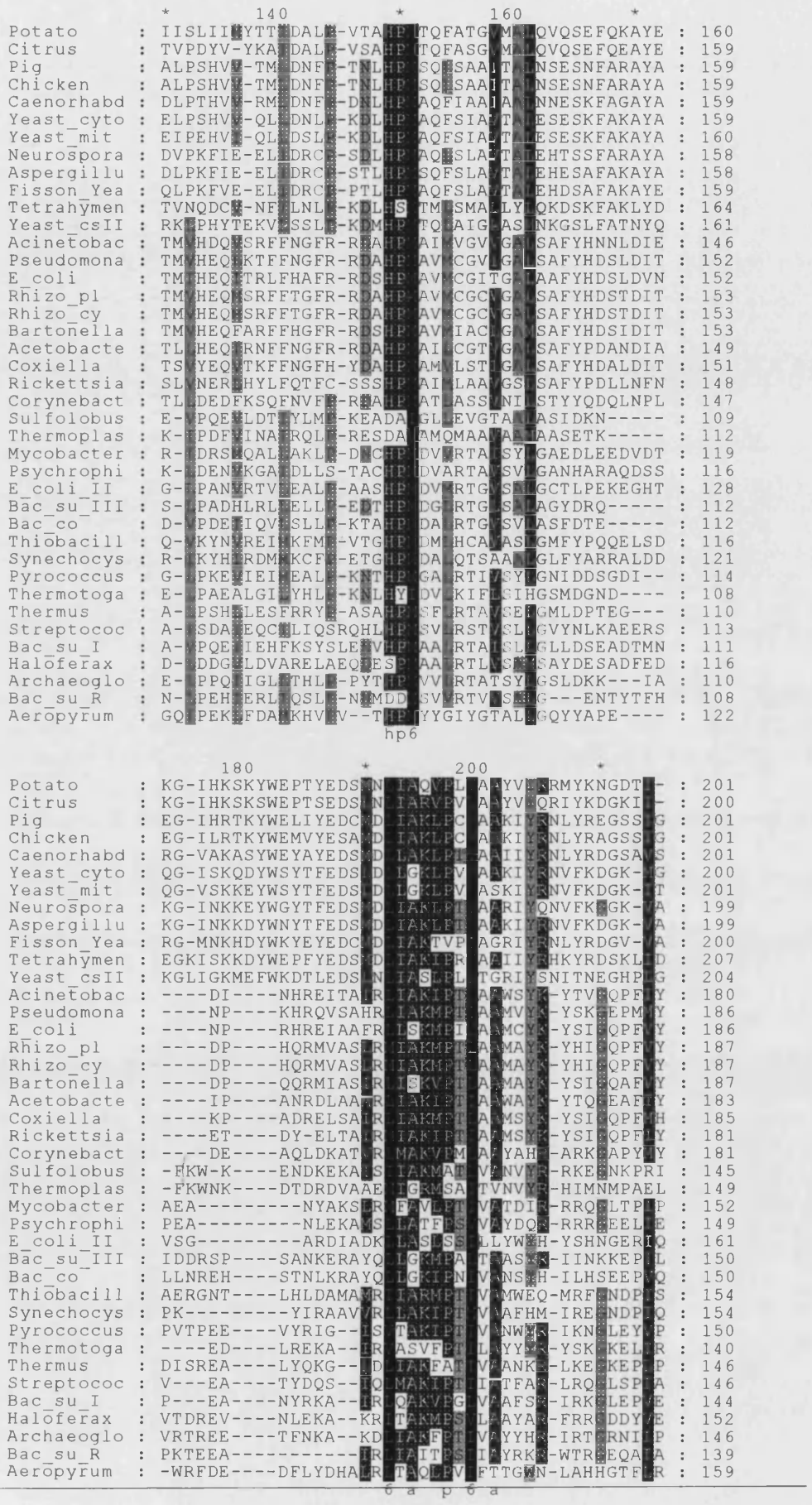
Saccharomyces cerevisiae (cii)	Yeast csII	Eukarya	Fungi
Saccharomyces cerevisiae (cyto)	Yeast cyto	Eukarya	Fungi
Saccharomyces cerevisiae (mit)	Yeast mit	Eukarya	Fungi
Saccharomyces pombe	Fission yea	Eukarya	Fungi
Solanum tuberosum	Potato	Eukarya	Plant
Sus scrofa	Pig	Eukarya	Animal
Tetrahymena thermophlia	Tetrahymen	Eukarya	Protozoa
Cyanobacteria anabaena	Cyanobacte	Eubacteria	Gram-ve
Archaeoglobus fulgidus	Archaeoglo	Archaea	Archaea
Aeropyrum pernix	Aeropyrum	Archaea	Gram-ve
Thermatoga sp	Thermotoga	Eubacteria	Gram-ve
Rickettsia sp	Rickettsia	Eubacteria	Gram-ve
Pseudomonas sp	Pseudomonas	Eubacteria	Gram-ve

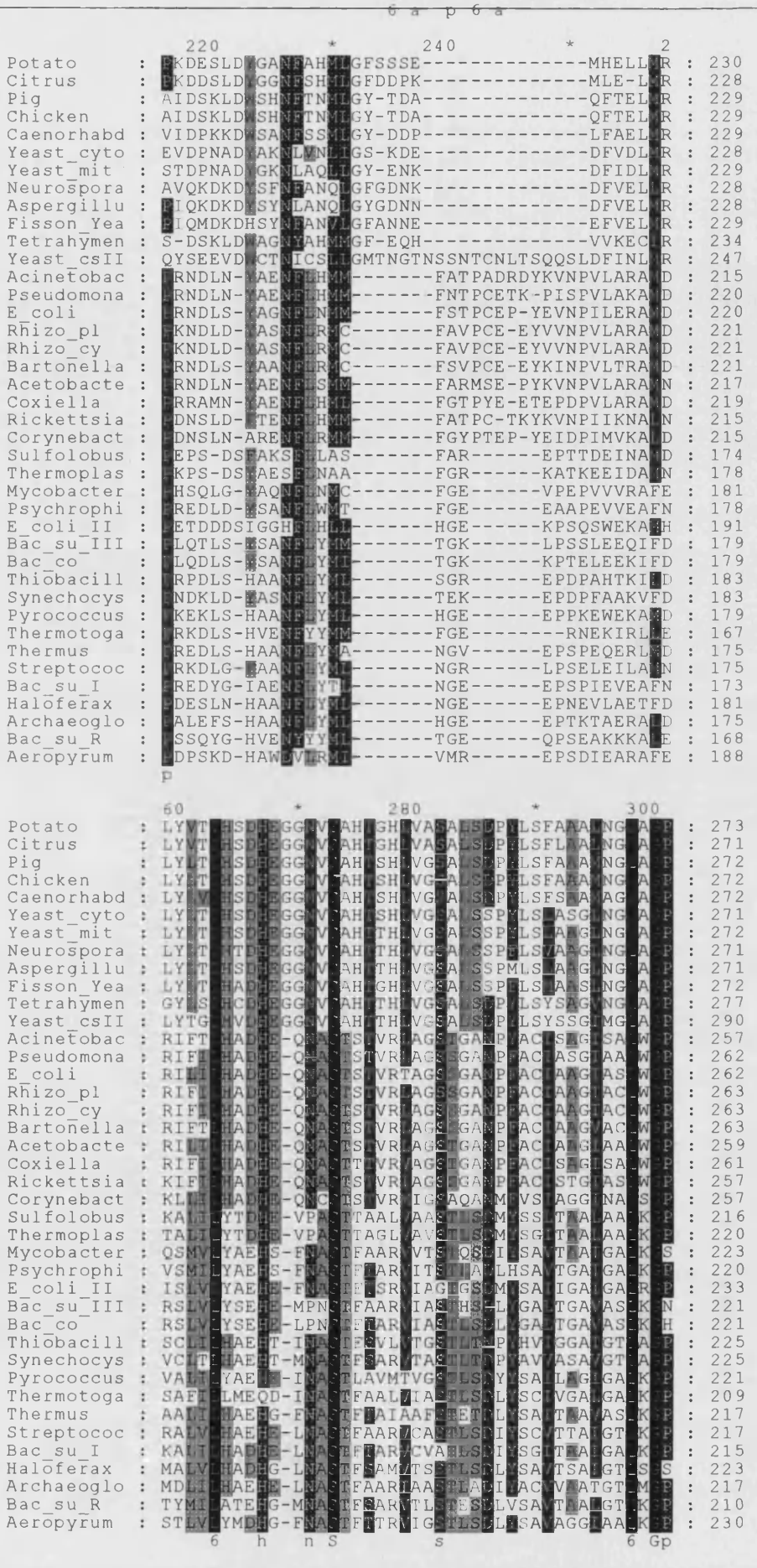
Appendix 3

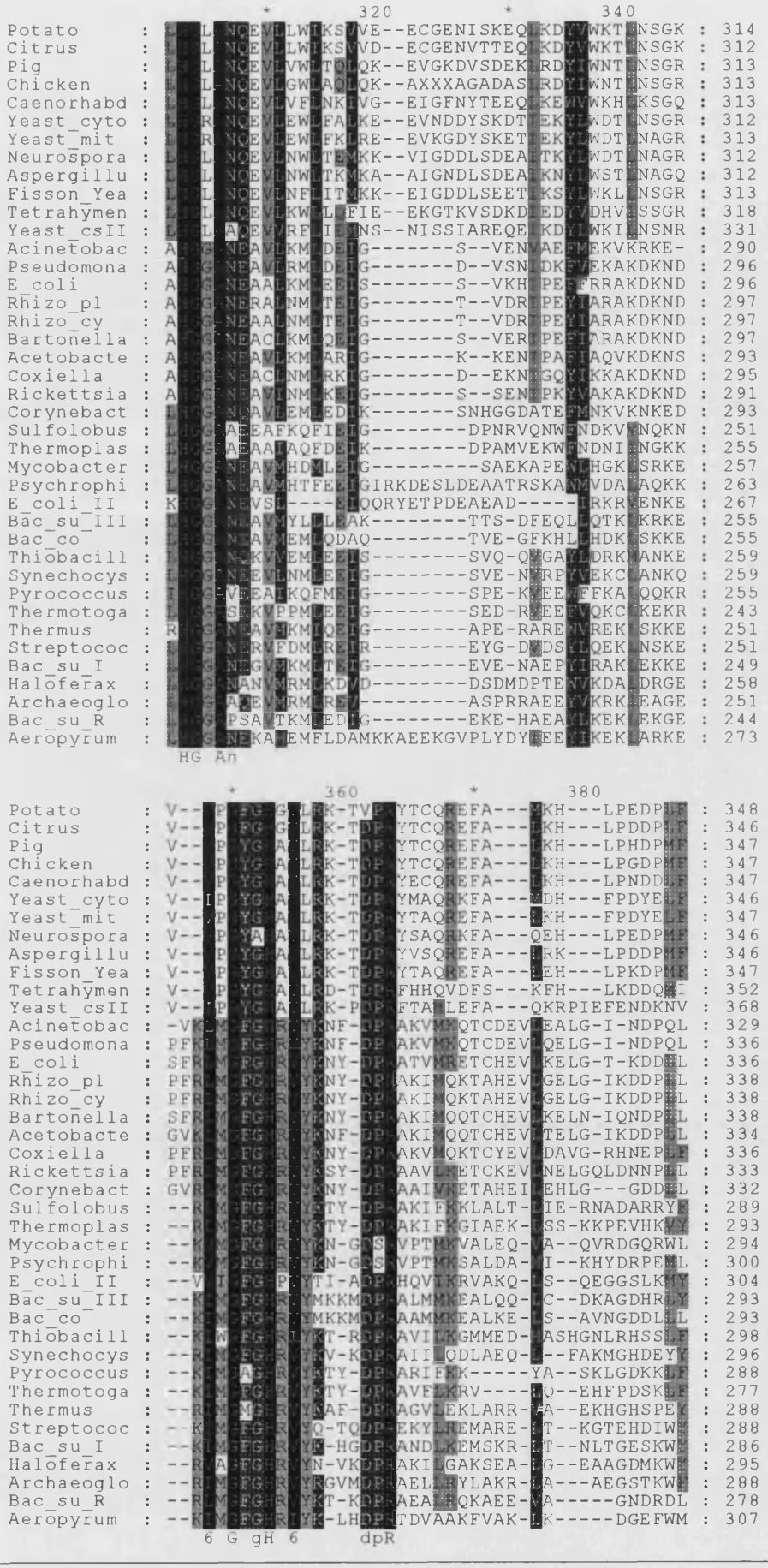
Multiple alignment of citrate synthase from the organisms used to produce the phylogenetic trees shown in chapter 4.

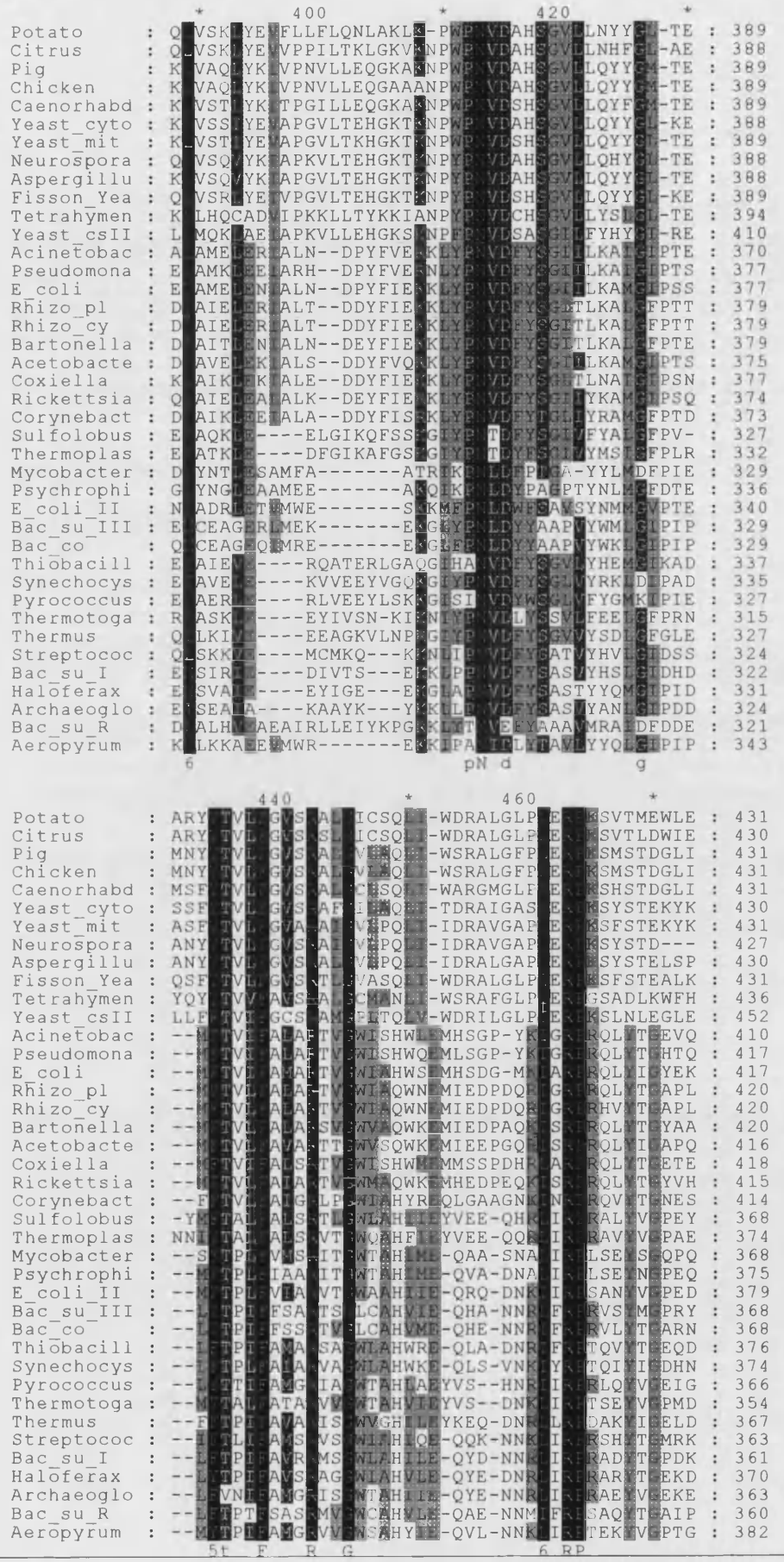
```
                    *         20          *         40
Potato      : ---SSGLDLRSELVQELIPEQQDRLKKIKSDM-KGSIGNITV- :  38
Citrus      : ---SADLDLHSQL-KEMIPEQQERLKKVKSDLGKAQLGNITI- :  38
Pig         : ---ASSTNLKDIL-ADLIPKEQARIKTFRQQHGNTVVGQITV- :  38
Chicken     : ---ASSTNLKDVL-AALIPKEQARIKTFRQQHGGTALGQITV- :  38
Caenorhabd  : ---EGSTNLKEVL-SKKIPAHNAKVKSFRTEHGSTVVQNVNI- :  38
Yeast_cyto  : ---SQEKTLKERF-SEIYPIHAQDVRQFVKEHGKTKISDVLL- :  38
Yeast_mit   : ---ASEQTLKERF-AEIIPAKAQEIKKFKKEHGKTVIGEVLLE :  39
Neurospora  : ----KTQTLKERF-AELLPENIEKIKALRKEHGSKVVDKVTL- :  37
Aspergillu  : ----KAKSLKETF-AEKLPAELEKVKKLRKEHGSKVIGEVTL- :  37
Fisson_Yea  : ---TRSSSLKDRL-AELIPEKQAEIKKFRAEHGQDVIGEVTI- :  38
Tetrahymen  : ----SQTNLKKVI-AEIIPQKQAELKEVKEKYGDKVVGQYTV- :  37
Yeast_csII  : ---SSALTLKEAL-ENVIPKKRDAVKKLKACYGSTFVGPITIS :  39
Acinetobac  : ------VLHLDGKEIELPIYSGTLGPDVIDVKDVLA-SGHFTF :  36
Pseudomona  : ADKKAQLIIEGSAPVELPVLSGTMGPDVVDVRGLTA-TGHFTF :  42
E_coli      : ADTKAKLTLNGDTAVELDVLKGTLGQDVIDIRTLGS-KGVFTF :  42
Rhizo_pl    : MDNNNACVLVDGHSAELKLRSSTIGPNVLGIGSLYEQTKMFTY :  43
Rhizo_cy    : MTEQSAKLTWGEKTVDLPVKTEPSAQVSLILVPFIRTPPLFTY :  43
Bartonella  : MSKNKAHITVNDKKIELSVRKGTLGPDVIEIASLYKETDTFTY :  43
Acetobacte  : ----TATISVDGKSAEMPVLSGTLGPDVIDIRKLPAQLGVFTF :  39
Coxiella    : -SNRKAKLSFENQSVEFPIYSPTLGKDVIDVKTL-GNHGAYAL :  41
Rickettsia  : -----AELKIRGKLFKLPILKASIGKDVIDISRVSAEADYFTY :  38
Corynebact  : ------VLHYPGGEFEMDIIEASEGNNGVVLGKMLSETGLITF :  37
Sulfolobus  : ------------------------------------------MSVV :   4
Thermoplas  : ------------------------------------------PETEEI :   6
Mycobacter  : ---------------------------------TTATESEAPRI :  11
Psychrophi  : ------------------------------------------MTEPTI :   6
E_coli_II   : ------------------------MSDTTILQNSTHVIKPKK :  18
Bac_su_III  : ------------------------------------------MEEKQHY :   7
Bac_co      : ------------------------------------------MVNTNQF :   7
Thiobacill  : ------------------------------------------MAEPNF :   6
Synechocys  : ------------------------------MNYMMTDNEVF :  11
Pyrococcus  : ------------------------------------------NTEKYL :   6
Thermotoga  : ------------------------------------------LI :   2
Thermus     : ------------------------------------------MEV :   3
Streptococ  : ------------------------------------------MAE :   3
Bac_su_I    : ------------------------------------------TA :   2
Haloferax   : ------------------------------------------MSGEL :   5
Archaeoglo  : -------------------------------------------M :   1
Bac_su_R    : -------------------------------------------V :   1
Aeropyrum   : --------------------------MSQQPECRVEGNYIIV :  16
```

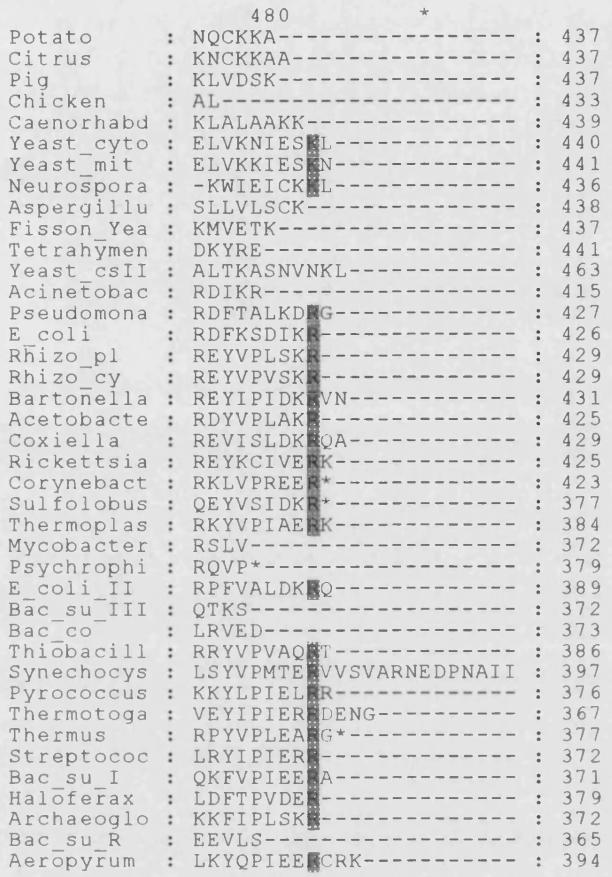



hp6
6a p6a


Potato
Citrus
Pig
Chicken
Caenorhabd
Yeast_cyto
Yeast_mit
Neurospora
Aspergillu
Fisson_Yea
Tetrahymen
Yeast_csII
Acinetobac
Pseudomona
E_coli
Rhizo_pl
Rhizo_cy
Bartonella
Acetobacte
Coxiella
Rickettsia
Corynebact
Sulfolobus
Thermoplas
Mycobacter
Psychrophi
E_coli_II
Bac_su_III
Bac_co
Thiobacill
Synechocys
Pyrococcus
Thermotoga
Thermus
Streptococ
Bac_su_I
Haloferax
Archaeoglo
Bac_su_R
Aeropyrum







Produced using PILEUP (GCG). Presentation and shading of conserved regions produced using GENEDOC. Residues highlighted in black indicate 100% conservation in this particular alignment, dark grey indicates 80% identity and light grey 60% identity. The full names of the organisms used in this alignment can be found in appendix 2.

Appendix 4

Crystallisation Conditions used that Led to Precipitation of the Protein

pH	Matrix Screen Conditions
8	70% MPD
7	70% MPD
6	40% MPD
6	60% MPD
6	70% MPD
6	35% Propan-2-ol
6	45% Propan-2-ol
7	25% Propan-2-ol
7	35% Propan-2-ol
7	45% Propan-2-ol
8	45% Propan-2-ol
4	0.8M Ammonium sulfate, 0.1M Citric Acid
4	3.2M Ammonium sulfate, 0.1M Citric Acid
4	3.2M Ammonium sulfate, 0.1M Citric Acid
4	2.4M Ammonium sulfate, 0.1M Citric Acid